

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES  
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum  
Internationales Büro



(43) Internationales Veröffentlichungsdatum  
4. Januar 2001 (04.01.2001)

PCT

(10) Internationale Veröffentlichungsnummer  
**WO 01/00802 A2**

(51) Internationale Patentklassifikation<sup>7</sup>: C12N 15/00

(21) Internationales Aktenzeichen: PCT/EP00/05853

(22) Internationales Anmeldedatum:  
23. Juni 2000 (23.06.2000)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:  
199 29 365.1 25. Juni 1999 (25.06.1999) DE

(71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von US): BASF-LYNX BIOSCIENCE AG [DE/DE]; D-69120 Heidelberg (DE).

(72) Erfinder; und

(75) Erfinder/Anmelder (nur für US): MACK, Matthias [DE/DE]; Mönchhofstrasse 3 C, D-69120 Heidelberg (DE). HERBSTER, Karin [DE/DE]; Kolpingstrasse 23a, D-76694 Forst (DE).

(74) Anwalt: GOLDSCHIED, Bettina; BASF Aktiengesellschaft, D-67056 Ludwigshafen (DE).

(81) Bestimmungsstaaten (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Bestimmungsstaaten (regional): ARIPO-Patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Veröffentlicht:**

— Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.

Zur Erklärung der Zweibuchstaben-Codes, und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

(54) Title: PARTIAL SEQUENCES OF THE GENES OF THE PRIMARY AND SECONDARY METABOLISM FROM *CORYNEBACTERIUM GLUTAMICUM* AND THEIR USE IN THE MICROBIAL PRODUCTION OF PRIMARY AND SECONDARY METABOLITES

(54) Bezeichnung: TEILSEQUENZEN DER GENE DES PRIMÄR- UND SEKUNDÄRMETABOLISMUS AUS *CORYNEBACTERIUM GLUTAMICUM* UND IHR EINSATZ ZUR MIKROBIELLEN HERSTELLUNG VON PRIMÄR- UND SEKUNDÄRMETABOLITEN

(57) Abstract: The invention relates to methods of producing primary and secondary metabolites using genetically engineered organisms.

(57) Zusammenfassung: Die vorliegende Erfindung befaßt sich mit Herstellungsverfahren für Primär- und Sekundärmetabolite mit Hilfe gentechnisch veränderter Organismen.

WO 01/00802 A2

Teilsequenzen der Gene des Primär- und Sekundärmetabolismus aus *Corynebacterium glutamicum* und ihr Einsatz zur mikrobiellen Herstellung von Primär- und Sekundärmetaboliten

5

#### Beschreibung

Die vorliegende Erfindung befaßt sich mit den Herstellungsverfahren für Primär- und Sekundärmetabolite mit Hilfe eines gentechnisch veränderten Organismus. Diese Erfindung besteht in Teilsequenzen von Genen, die anabolische und katabolische Enzyme aus *Corynebacterium glutamicum* kodieren, und aus ihrem Einsatz zur mikrobiellen Herstellung von Metaboliten.

15 Die Konzentrationen der Metabolite sind in lebenden Zellen gewöhnlich gut ausbalanciert und überschreiten nicht eine gewisse Grenze. Unter manchen Wachstumsbedingungen oder als Folge einer gentechnischen Veränderung können sie allerdings im Überschuß gebildet und in das Kulturmedium ausgeschieden werden. Für das Zellwachstum kann man relativ billige Stoffe als Kohlenstoffquelle verwenden. Mit Hilfe des biochemischen Potentials der Zellen (in den meisten Fällen mikrobiellen Ursprungs) oder der Enzyme lassen sich diese preiswerten Stoffe in ein breites Spektrum wertvollere Substanzen umwandeln. Zur fermentativen Herstellung von Metaboliten zu Verkaufszwecken setzt man insbesondere Mikroorganismen ein. Mikroorganismen lassen sich durch gentechnische Veränderung der Biosynthesewege in ihrer Biosyntheseleistung auf bestimmte Metabolite hin optimieren, und man erzielt dadurch höhere Syntheseleistungen. Gentechnische Veränderung meint hier, daß die Anzahl der Kopien oder die Geschwindigkeit der Transkription bestimmter Gene für bestimmte Synthesewege erhöht ist. Allerdings muß man die geeigneten Zielgene für diese Verbesserung zuerst identifizieren. Wir beschreiben nun im folgenden die Zielgene und Teilsequenzen davon, die durch Klonen der DNA und anschließende Sequenzierung mit dem Ziel der Stammverbesserung identifiziert wurden.

Ein Teil der Erfindung besteht in einem Genfragment mit einer Nucleotidsequenz, die in der SEQ ID NR. 1 beschrieben ist oder sich von dieser Sequenz SEQ ID NR. 1 durch Substitution, Insertion oder Deletion von bis zu 20 % der Nucleotide ableitet.

Ein weiterer Teil der Erfindung besteht in einem Genfragment mit einer Nucleotidsequenz, die in der SEQ ID NR. 2 beschrieben ist oder sich von dieser Sequenz SEQ ID NR. 2 durch Substitution, Insertion oder Deletion von bis zu 20 % der Nucleotide ableitet.

## 2

Ein weiterer Teil der Erfindung besteht in einem Genfragment mit einer Nucleotidsequenz, die in der SEQ ID NR. 3 beschrieben ist oder sich von dieser Sequenz SEQ ID NR. 3 durch Substitution, Insertion oder Deletion von bis zu 20 % der Nucleotide ableitet.

5

Ein weiterer Teil der Erfindung besteht in einem Genfragment mit einer Nucleotidsequenz, die in der SEQ ID NR. 4 beschrieben ist oder sich von dieser Sequenz SEQ ID NR. 4 durch Substitution, Insertion oder Deletion von bis zu 20 % der Nucleotide ableitet.

10

Ein weiterer Teil der Erfindung besteht in einem Genfragment mit einer Nucleotidsequenz, die in der SEQ ID NR. 5 beschrieben ist oder sich von dieser Sequenz SEQ ID NR. 5 durch Substitution, Insertion oder Deletion von bis zu 20 % der Nucleotide ableitet.

15

Ein weiterer Teil der Erfindung besteht in einem Genfragment mit einer Nucleotidsequenz, die in der SEQ ID NR. 6 beschrieben ist oder sich von dieser Sequenz SEQ ID NR. 6 durch Substitution, Insertion oder Deletion von bis zu 20 % der Nucleotide ableitet.

20

Ein weiterer Teil der Erfindung besteht in einem Genfragment mit einer Nucleotidsequenz, die in der SEQ ID NR. 7 beschrieben ist oder sich von dieser Sequenz SEQ ID NR. 7 durch Substitution, Insertion oder Deletion von bis zu 20 % der Nucleotide ableitet.

25

Ein weiterer Teil der Erfindung besteht in einem Genfragment mit einer Nucleotidsequenz, die in der SEQ ID NR. 8 beschrieben ist oder sich von dieser Sequenz SEQ ID NR. 8 durch Substitution, Insertion oder Deletion von bis zu 20 % der Nucleotide ableitet.

30

Ein weiterer Teil der Erfindung besteht in einem Genfragment mit einer Nucleotidsequenz, die in der SEQ ID NR. 9 beschrieben ist oder sich von dieser Sequenz SEQ ID NR. 9 durch Substitution, Insertion oder Deletion von bis zu 20 % der Nucleotide ableitet.

35

Ein weiterer Teil der Erfindung besteht in einem Genfragment mit einer Nucleotidsequenz, die in der SEQ ID NR. 10 beschrieben ist oder sich von dieser Sequenz SEQ ID NR. 10 durch Substitution, Insertion oder Deletion von bis zu 20 % der Nucleotide ableitet.

40

Ein weiterer Teil der Erfindung besteht in einem Genfragment mit einer Nucleotidsequenz, die in der SEQ ID NR. 11 beschrieben ist oder sich von dieser Sequenz SEQ ID NR. 11 durch Substitution, Insertion oder Deletion von bis zu 20 % der Nucleotide ableitet.

45

## 3

Ein weiterer Teil der Erfindung besteht im Einsatz der Nucleotidsequenz SEQ ID NR. 1 oder SEQ ID NR. 2 oder SEQ ID NR. 3 oder SEQ ID NR. 4 oder SEQ ID NR. 5 oder SEQ ID NR. 6 oder SEQ ID NR. 7 oder SEQ ID NR. 8 oder SEQ ID NR. 9 oder SEQ ID NR. 10 oder SEQ ID NR. 11 zur Konstruktion genetisch modifizierter Mikroorganismen.

Die vollständigen Gene lassen sich mit Hilfe konventioneller Techniken wie Hybridisierung herstellen, wobei man von den oben offenbarten Genfragmenten ausgeht. Diese Gene lassen sich einsetzen zur Konstruktion rekombinanter Wirtsorganismen, die die Biosynthese wertvoller Bioprodukte, wie Aminosäuren, Fettsäuren, Kohlenhydrate, Vitamine und Kofaktoren ermöglichen. Die biologische Aktivität dieser Gene wird im experimentellen Teil dieser Beschreibung offenbart. Mit Hilfe dieser Gene wird es möglich, Engpässe bei der Biosynthese von Bioprodukten zu umgehen und so die Syntheseleistung mikrobieller Systeme zu steigern.

Ein weiterer Gesichtspunkt dieser Erfindung besteht in einem Expressions-Vektor mit zumindest einem der oben erwähnten Polynucleotide. Der Expressions-Vektor verbindet funktionell eines oder mehrere dieser Polynucleotide mit regulatorischen Einheiten wie Promotoren, Terminatoren, ribosomale Bindungsstellen und dergleichen. Gewöhnlich gehören zu einem Expressions-Vektor weitere Einheiten wie Genmarker und Replikationsabschnitte.

Ein weiterer Gesichtspunkt der Erfindung besteht in der mit einem Expressions-Vektor transformierten Wirtszelle.

Zur gentechnischen Veränderung kann man jeden beliebigen prokaryontischen Mikroorganismus verwenden, vorzugsweise *Corynebacterium*- und *Bacillus*-Arten, aber auch jeden beliebigen eukaryontischen Mikroorganismus, vorzugsweise Hefestämme der Gattung *Ashbya*, *Candida*, *Pichia*, *Saccharomyces* und *Hansenula*.

Ein weiterer Gesichtspunkt der Erfindung besteht in einer Methode zur Herstellung und Reinigung eines Polypeptids, die in folgenden Schritten besteht:

- (a) Kultivierung der Wirtszelle aus Anspruch 3 unter Bedingungen, die für die Expression des Peptids geeignet sind; und
- (b) Gewinnung des Polypeptids aus der Wirtszellkultur.

In den folgenden Beispielen wird die Erfindung detaillierter beschrieben.

## Beispiel 1

Herstellung einer Genombibliothek von *Corynebacterium glutamicum* ATCC 13032

5

Die DNA aus dem Genom von *Corynebacterium glutamicum* ATCC 13032 läßt sich nach Standardmethoden gewinnen, die z.B. von Altenbuchner, J. und Cullum, J. (1984, Mol. Gen. Genet. 195:134-138) beschrieben sind. Die Genom-Bibliothek läßt sich daraus mit jedem beliebigen Klonierungsvektor, z.B. pBluescript II KS- (Stratagene) oder ZAP Express<sup>TM</sup> (Stratagene), nach Standardvorschriften gewinnen (z.B. Sambrook, J. et al. (1989) Molecular cloning: a laboratory manual, Cold Spring Harbor Laboratory Press). Jede beliebige Fragmentgröße kann man dabei verwenden, vorzugsweise, Sau3AI-Fragmente mit einer Länge von 1 kb, die sich in Klonierungsvektoren mit verdautem BamHI einbinden lassen.

## Beispiel 2

## 20 Analyse der Nucleinsäuresequenzen der Genombibliothek

Aus der im Beispiel 1 hergestellten Genombibliothek kann man einzelne *E. coli*-Klone auswählen. *E. coli*-Zellen werden nach Standardmethoden in geeigneten Medien kultiviert (z.B. LB ergänzt mit 100 mg/l Ampicillin), und danach läßt sich dann die Plasmid-DNA isolieren. Klont man Genomfragmente aus der DNA von *Corynebacterium glutamicum* in pBluescript II KS- (siehe Beispiel 1), läßt sich die DNA mit Hilfe der Oligonucleotide 5'-AATTAAC-CCTCACTAAAGGG-3' und 5'-GTAATACGACTCACTATAGGGC-3' sequenzieren.

30

## Beispiel 3

Computeranalyse der isolierten Nucleinsäuresequenzen

35 Die Nucleotidsequenzen lassen sich z.B. mit Hilfe des BLASTX-Algorithmus (Altschul et al. (1990) J. Mol. Biol. 215: 403-410) aneinanderfügen. Auf diesem Weg kann man neuartige Sequenzen entdecken und die Funktion dieser neuartigen Gene aufklären.

## 40 Beispiel 4

Identifizierung eines *E. coli*-Klons mit einem Genfragment für die Fettsäuresynthase (2.3.1.85)

45 Bei der Analyse der *E. coli*-Klone, wie sie im Beispiel 2 beschrieben wurde, an die sich die im Beispiel 3 beschriebene Analyse der dabei erhaltenen Sequenzen anschloß, fand sich eine

## 5

Sequenz, wie sie mit SEQ ID NR. 1 beschrieben ist. Bei der Anwendung des BLASTX-Algorithmus (siehe Beispiel 3) ergab diese Sequenz Ähnlichkeit mit Fettsäuresynthasen aus verschiedenen Organismen. Die größte Ähnlichkeit war mit einem Fragment mit 519 Basenpaaren für die Fettsäuresynthase aus *Corynebacterium ammoniagenes* gegeben (NRDB Q04846; 68% Übereinstimmung auf der Stufe der Aminosäuren).

## Beispiel 5

10

Identifizierung eines *E. coli*-Klons mit dem Gen für die Phytoen-Dehydrogenase (EC 1.3.-.-)

Bei der Analyse der *E. coli*-Klone, wie sie im Beispiel 2 beschrieben wurde und an die sich die im Beispiel 3 beschriebene Analyse der dabei erhaltenen Sequenzen anschloß, fand sich eine Sequenz, die als SEQ ID NR. 2 beschrieben ist. Bei der Anwendung des BLASTX-Algorithmus (siehe Beispiel 3) zeigte diese Sequenz Ähnlichkeit mit Phytoen-Dehydrogenasen aus verschiedenen Organismen. Die größte Ähnlichkeit ergab sich mit der Phytoen-Dehydrogenase aus *Methanobacterium thermoautotrophicum* (NRDB O27835; 37% Übereinstimmung auf der Stufe der Aminosäuren).

## Beispiel 6

25

Identifizierung eines *E. coli*-Klons mit dem Gen für die Alkohol-Dehydrogenase (EC 1.1.1.1)

Bei der Analyse der *E. coli*-Klone, wie sie im Beispiel 2 beschrieben wurde und an die sich die im Beispiel 3 beschriebene Analyse der dabei erhaltenen Sequenzen anschloß, fand sich eine Sequenz, die als SEQ ID NR. 3 beschrieben ist. Bei der Anwendung des BLASTX-Algorithmus (siehe Beispiel 3) zeigte diese Sequenz Ähnlichkeit mit Alkohol-Dehydrogenasen aus verschiedenen Organismen. Die größte Ähnlichkeit ergab sich mit der Alkohol-Dehydrogenase aus *Bacillus stearothermophilus* (NRDB P42327; 50% Übereinstimmung auf der Stufe der Aminosäuren).

## Beispiel 7

40

Identifizierung eines *E. coli*-Klons mit einem Genfragment für ein Homologes der Adenosylmethionin-8-Amino-7-oxononanoat-Aminotransferase (EC 2.6.1.62)

Bei der Analyse der *E. coli*-Klone, wie sie im Beispiel 2 beschrieben wurde und an die sich die im Beispiel 3 beschriebene Analyse der dabei erhaltenen Sequenzen anschloß, fand sich eine



## 6

Sequenz, die als SEQ ID NR. 4 beschrieben ist. Bei der Anwendung des BLASTX-Algorithmus (siehe Beispiel 3) zeigte diese Sequenz Ähnlichkeit mit Adenosylmethionin-8-Amino-7-oxononanoat-Aminotransferasen aus verschiedenen Organismen. Die größte Ähnlichkeit  
5 ergab sich mit einem aus 342 Basenpaaren bestehenden Fragment für die Adenosylmethionin-8-amino-7-oxononanoat-Aminotransferase aus *Erwinia herbicola* (NRDB P53656; 40% Übereinstimmung auf der Stufe der Aminosäuren).

## 10 Beispiel 8

Identifizierung eines *E. coli*-Klons mit einem Genfragment für ein Homologes der Phosphoglycerat-Mutase 2 (EC 5.4.2.1)

15 Bei der Analyse der *E. coli*-Klone, wie sie im Beispiel 2 beschrieben wurde und an die sich die im Beispiel 3 beschriebene Analyse der dabei erhaltenen Sequenzen anschloß, fand sich eine Sequenz, die als SEQ ID NR. 5 beschrieben ist. Bei der Anwendung des BLASTX-Algorithmus (siehe Beispiel 3) zeigte diese Sequenz  
20 Ähnlichkeit mit Phosphoglycerat-Mutasen 2 aus verschiedenen Organismen. Die größte Ähnlichkeit ergab sich mit einem aus 204 Basenpaaren bestehenden Fragment für die Phosphoglycerat-Mutase 2 aus *Mycobacterium tuberculosis* (NRDB P71724; 54% Übereinstimmung auf der Stufe der Aminosäuren).

25

## Beispiel 9

Identifizierung eines *E. coli*-Klons mit einem Genfragment für die Xylulose-Kinase (EC 2.7.1.17)

30

Bei der Analyse der *E. coli*-Klone, wie sie im Beispiel 2 beschrieben wurde und an die sich die im Beispiel 3 beschriebene Analyse der dabei erhaltenen Sequenzen anschloß, fand sich eine Sequenz, die als SEQ ID NR. 6 beschrieben ist. Bei der Anwendung  
35 des BLASTX-Algorithmus (siehe Beispiel 3) zeigte diese Sequenz Ähnlichkeit mit Xylulose-Kinasen aus verschiedenen Organismen. Die größte Ähnlichkeit ergab sich mit einem aus 633 Basenpaaren bestehenden Fragment für die Xylulose-Kinase aus *Streptomyces rubiginosus* (NRDB P27156; 48% Übereinstimmung auf der Stufe der  
40 Aminosäuren).

## Beispiel 10

Identifizierung eines *E. coli*-Klons mit einem Genfragment für  
45 eine Fettsäure-CoA-Ligase für langkettige Fettsäuren (EC 6.2.1.3)

## 7

Bei der Analyse der *E. coli*-Klone, wie sie im Beispiel 2 beschrieben wurde und an die sich die im Beispiel 3 beschriebene Analyse der dabei erhaltenen Sequenzen anschloß, fand sich eine Sequenz, die als SEQ ID NR. 7 beschrieben ist. Bei der Anwendung  
5 des BLASTX-Algorithmus (siehe Beispiel 3) zeigte diese Sequenz Ähnlichkeit mit Fettsäure-CoA-Ligasen für langkettige Fettsäuren aus verschiedenen Organismen. Die größte Ähnlichkeit ergab sich mit einem aus 369 Basenpaaren bestehenden Fragment für die Fettsäure-CoA-Ligase für langkettige Fettsäuren aus *Archaeoglobus*  
10 *fulgidus* (NRDB 030302; 48% Übereinstimmung auf der Stufe der Aminosäuren).

## Beispiel 11

15 Identifizierung eines *E. coli*-Klons mit einem Genfragment für die Guanosinpentaphosphat-Synthetase

Bei der Analyse der *E. coli*-Klone, wie sie im Beispiel 2 beschrieben wurde und an die sich die im Beispiel 3 beschriebene  
20 Analyse der dabei erhaltenen Sequenzen anschloß, fand sich eine Sequenz, die als SEQ ID NR. 8 beschrieben ist. Bei der Anwendung des BLASTX-Algorithmus (siehe Beispiel 3) zeigte diese Sequenz Ähnlichkeit mit Guanosinpentaphosphat-Synthetasen aus verschiedenen Organismen. Die größte Ähnlichkeit ergab sich mit einem aus  
25 606 Basenpaaren bestehenden Fragment für die Guanosinpentaphosphate-Synthetase aus *Streptomyces coelicolor* (NRDB 086656; 70% Übereinstimmung auf der Stufe der Aminosäuren).

## Beispiel 12

30 Identifizierung eines *E. coli*-Klons mit einem Genfragment für ein NTRB-Homologes

Bei der Analyse der *E. coli*-Klone, wie sie im Beispiel 2 beschrieben wurde und an die sich die im Beispiel 3 beschriebene  
35 Analyse der dabei erhaltenen Sequenzen anschloß, fand sich eine Sequenz, die als SEQ ID NR. 9 beschrieben ist. Bei der Anwendung des BLASTX-Algorithmus (siehe Beispiel 3) zeigte diese Sequenz Ähnlichkeit mit NTRB-Homologen aus verschiedenen Organismen. NTRB  
40 ist ein Regulatorgen für die Transkription, das an der Regulierung der Stickstoffassimilation beteiligt ist. Die größte Ähnlichkeit ergab sich mit einem aus 645 Basenpaaren bestehenden Fragment für NTRB aus *Mycobacterium leprae* (NRDB Q50049; 61% Übereinstimmung auf der Stufe der Aminosäuren).

45



## Beispiel 13

Identifizierung eines *E. coli*-Klons, der ein *nifS*-Homologes enthält

5

Bei der Analyse der *E. coli*-Klone, wie sie im Beispiel 2 beschrieben wurde und an die sich die im Beispiel 3 beschriebene Analyse der dabei erhaltenen Sequenzen anschloß, fand sich eine Sequenz, die als SEQ ID NR. 10 beschrieben ist. Bei der Anwendung  
10 des BLASTX-Algorithmus (siehe Beispiel 3) zeigte diese Sequenz Ähnlichkeit mit *nifS* aus verschiedenen Organismen. *NifS* ist an der Stickstofffixierung beteiligt. Die größte Ähnlichkeit ergab sich mit einem aus 594 Basenpaaren bestehenden Fragment für *nifS* aus *Mycobacterium leprae* (NRDB Q49690; 62% Übereinstimmung auf  
15 der Stufe der Aminosäuren).

## Beispiel 14

Identifizierung eines *E. coli*-Klons, der ein *nifU*-Homologes ent-  
20 hält

Bei der Analyse der *E. coli*-Klone, wie sie im Beispiel 2 beschrieben wurde und an die sich die im Beispiel 3 beschriebene Analyse der dabei erhaltenen Sequenzen anschloß, fand sich eine  
25 Sequenz, die als SEQ ID NR. 11 beschrieben ist. Bei der Anwendung des BLASTX-Algorithmus (siehe Beispiel 3) zeigte diese Sequenz Ähnlichkeit mit *nifU* aus verschiedenen Organismen. *NifU* ist an der Stickstofffixierung beteiligt. Die größte Ähnlichkeit ergab sich mit einem aus 339 Basenpaaren bestehenden Fragment für *nifU*  
30 aus *Mycobacterium leprae* (NRDB Q49683; 61% Übereinstimmung auf der Stufe der Aminosäuren).

## Sequenzliste

## 35 (I) Allgemeine Angaben

## (1) Anmelder:

(A) Name:	BASF-LYNX Bioscience AG
40 (B) Straße:	Im Neuenheimer Feld 515
(C) Stadt:	Heidelberg
(D) Land:	Deutschland
(E) Postleitzahl:	69120
(F) Telephon:	06221/4546
45 (G) Telefax:	06221/454770

9

(2) Titel: Sequenzen der Gene für den Primär- und Sekundärmetabolismus im *Corynebacterium glutamicum* und ihr Einsatz zur mikrobiellen Herstellung von Primär- und Sekundärmetaboliten

5

(3) Anzahl der Sequenzen: 11

(4) Art der mit dem Computer lesbaren Form:

10

(A) Datenträger: Diskette  
(B) Computer: IBM PC kompatibel  
(C) Betriebssystem: Windows NT  
(D) Software: Microsoft®word 97 SR-1

15

(I) Angaben zur SEQ ID NR. 1:

(1) Sequenzcharakteristika:

20 (A) Länge: 693 Basenpaare  
(B) Art: Nucleinsäure  
(C) Strangtyp: Doppelstrang  
(D) Topologie: linear

25 (2) Molekülart: DNA  
(3) hypothetisch: nein  
(4) Antisense: nein

(5) Herkunft:

30

(A) Organismus: *Corynebacterium glutamicum*

(6) Beschreibung der Sequenz: SEQ ID NR. 1:

35 CTGTTNCCCGGGGATCAGATTCACNGGGTCNGCCAGTGAAGTCGACGGTGATTGGCGCGGATGC  
TGCCTGCTCGCGAACAGTGGAAGTTGCCTGGGACAGCAGTTTCTCTGCAATTTCTTGGGTGGAGT  
AGGTTTCCACGCCTGCTTCTTCAGCTGCCTTGACCAAAGGATCGTTGCCGCCCATGAGGCCGGTG  
CCGCGAACCCAACCGATGTGAGCGTGCACGAGGGAGGTGTGTGCTCCCCATGCAGCTTGCTCTGC  
GTTCCAACGGGTAACCACGGCGTCGAGAGCTGCCTTGATTACCGTATGCACCATCGCCACCGA  
40 AGCGTCCACGGTTTGGTGAACCTGGGATGACCACGTGCAGGCGGTGACCCACGTTGATGGAGGAG  
CCCAATGGCGCAAGACCTGCGATGAGGCGCTCAACAGACCAGAGCAGAAGTCGCATCTGGGATTC  
TGCTGTGGCCTGCATCTGCATGGATCCGGACACGCGAGGTGCCGCGAATGGGAACAGCAAGGTAN  
GGACCAAACCTGGCTTGACCAGCTTGGATGCNCGTGACGGNGGTGGCTGTGATCCACCCAGTTGA  
TGATGGCTCGATGTCTGATANGACTAAGTTACCGCACGATCACAGTGCTGCCTGCCGNTGCGGAA  
45 CGTCCTANNANTCTTGAGAATTCAGCCGNCTGGCCGAGTTGAN

10

(I) Angaben zu SEQ ID NR. 2:

(1) Sequenzcharakteristika:

- 5 (A) Länge: 1869 Basenpaare  
(B) Typ: Nucleinsäure  
(C) Strangtyp: Doppelstrang  
(D) Topologie: linear  
(2) Molekülart: DNA  
10 (3) hypothetisch: nein  
(4) Antisense: nein  
(5) Herkunft:
- 15 (B) Organismus: *Corynebacterium glutamicum*

(6) Beschreibung der Sequenz: SEQ ID NR. 2:

ACAATCAATGTCATGACCAGCGGTGATCCAAAATTAATTACGCTTCGTTCCTCCGCAATAACAAAGT  
20 TGTCGAAATAACAGATGCCCCGTGATCGTGCTTTTCATTAAGCAGGCTGATGAGCTGATCGTGAAGA  
TCGACAAACTGCTTGAGAGCAAAAGAAAAAGTCCCGTTAAGTCCGATCCACACTGTTGTCCCCG  
CGGAGACGCTTGAGCACGTTTTCTGCAGAGATCAAACACATAGATACGCCCACCCCTGGAAGTGT  
GGTGTACCTGCGTCATACAGGCCATCTACTTTGCGGGATTTGTTAGAACCCTAAAGAACGCCG  
ATTGTGCCAGGGTGTGTGAGGGGCCAATGGACCCGCCGCTCCAGGAGTTGTATCGGTCTGCGAAG  
25 TCGGCAGGGCCGATGGTGCCTGTCACAACAATGCGGCTTTCCAAACCATCAATGCCAGCCCATCG  
CCCAATTTGAGCCACTGCTGCTATTGCGATCCGGCCCACCATGTCAGATTCTTCTCCGTAAGCGG  
ACCCGTGACCAATGGAGACATCGGCGGGTACTGGGACCAGGATGAAGAGGTTCTCGTGGCCTTCG  
GGTGCAGCATCGGAATCTGTTGCGGAGGTCTTGGAGATCTAGATGGATTCTGAAGCCGGAATTC  
TGGGGTGGAGCCGTCGAAAACCTTTGCGGAAATCTTCGTCCCAGTCGGAGGAAAAAGCAGGGTGTG  
30 CTCCCCCTTCACGCCTGCCAAAACCAGCACAGTACTGAGGCCGGGTGTTTGTCTTCCAGCTCG  
TCTCCGGCTTCGCGCACAAACGAAGCAGGTAGGAGTTGGGTTTCGGTGTGGTGCTGATCAGCGCAG  
CTGATCACGATATCGGCTTCGATGAACTCTGAGCCGACTTGGACGCCTGTGGCGTTTCGGCCTTG  
GGTGGTGATTTGCGCTGACGGGGGTGCCGAGGTGGAGGACGGCGTCGTCGATAAGCGAAATTAGTG  
CCTTGATGAAGGCGGTGAAGCCGCCTCGGGGATAGGAGACGCCTTGGACGAGGTCGGTGTGGCTC  
35 ATGAGGTGATAGAGCGCCGGGTGTGCGAAGGGTCTGAGGAGAGGAAAACCTGCGGGGTAGCTTAA  
GATTTGGCGCAGTTTTGTATCGCGGAATTGGGTGTTGACCTTGACTTTTAGCGAGGTCGACAGGC  
TTGCTAGAAGTTTGGGTAAAAGGCGCAGCATGCCGGGGCTTAAGTATGGGATGAAGTTGGTGAAG  
TTGGTGTAGAGGAAGCCGTCGATGGCCAGGTTGTAGACCTGTGTGGCGGAGTCGATATAGGTGCG  
CAGTTTGGCGCCGGCGCCGGGTTCGCGGGATTGAAAAGCTCGGCCATCGCATCGATGTCGGAGG  
40 TGACGTCGATGAATTCGCCGTGGTTCGTCGATGACGCGGTAGGCGGGTTCAAGTGGCACGAGGTCG  
AGGTGGTCGTCGATGGAGGTGCCGCAGAGCTTAAAGAAGTGGGACATGGCGTCGGGCATGAGGTA  
CCAGCTGGGGCCGGTGTCCCAGCGGAAGCCGTCGAGTTTCAAGGTCCCGCGCGGCGCCGAGGT  
GCTCGTTTTGTTCGACGAGGTGGACTTCATATCCTTCGCGTAAGAGCAGTGCGGTGGTGGCTAGT  
CCTGCTAGTCCCCCGCCGATGACCCTGCTTTTGTGATTTTTTGAAACACTTCTTTCCACATTGCT  
45 TTGGTTGCCAGGCTGGCTTTTTTTCATAGACGGCACCCGAATCCGCCCGTTTTTTAAGTCTTCGAG

## 11

GGACGCGGATTCCAGGTTGTCCACGAGGCAACCGTAGAGATCGGTCGCGGGCGCGCACACCGGTTC  
GCGCGCCAAATGGCAGCAGCGGAATGCTCAGCCGGGCGGCATCCAAATC

(I) Angaben zur SEQ ID NR. 3:

5

(1) Sequenzcharakteristika:

(A) Länge: 1035 Basenpaare  
(B) Typ: Nucleinsäure  
10 (C) Strangtyp: Doppelstrang  
(D) Topologie: linear

(2) Molekülart: DNA  
(3) hypothetisch: nein  
15 (4) Antisense: nein

(5) Herkunft:

(C) Organismus: *Corynebacterium glutamicum*

20

(6) Beschreibung der Sequenz: SEQ ID NR. 3:

ATGACCACTGCTGCACCCCAAGAATTTACCGCTGCTGTTGTTGAAAAATTCGTTTCATGACGTGAC  
CGTGAAGGATATTGACCTTCCAAAGCCAGGGCCACACCAGGCATTGGTGAAGGTACTCACCTCCG,  
25 GCATTTGCCACACCGACCTCCACGCCTTGGAGGGCGATTGGCCAGTAAAGCCGGAACCACCATTC  
GTACCAGGACACGAAGGTGTAGGTGAAGTTGTTGAGCTCGGACCAGGTGAACACGATGTGAAGGT  
CGGCGATATTGTCGGCAATGCGTGGCTCTGGTCAGCGTGTGGCACCTGCGAATACTGCATCACCG  
GCAGGGAAACTCAGTGCAACGAAGCTGAGTATGGTGGCTACACCCAAAATGGATCCTTCGGCCAG  
TACATGCTGGTGGATACCCGTTACGCCGCTCGCATCCCAGACGGCGTGGACTACCTCGAAGCAGC  
30 ACCAATTCTGTGTGCAGGCGTGACTGTCTACAAGGCACTCAAAGTCTCTGAAACCCGCCCCGGGCC  
AATTCATGGTGATCTCCGGTGTGCGCGGACTTGGCCACATCGCAGTCCAATACGCAGCGGCGATG  
GGCATGCGTGTGATTGCGGTAGATATTGCCGATGACAAGCTGGAACCTTGCCCGTAAGCACGGTGC  
GGAATTTACCGTGAATGCGCGTAATGAAGATTCAGGCGAAGCTGTACAGAAGTACACCAACGGTG  
GCGCACACGGCGTGCTTGTGACTGCAGTTCACGAGGCAGCATTCGGCCAGGCACTGGATATGGCT  
35 CGACGTGCAGGAACAATTGTGTTCAACGGTCTGCCACCGGGAGAGTTCCCAGCATCCGTGTTCAA  
CATCGTATTCAAGGGCCTGACCATCCGTGGATCCCTCGTGGGAACCCGCCAAGACTTGGCCGAAG  
CGCTCGATTTCTTTGCACGCGGACTAATCAAGCCAACCGTGAGTGAGTGCTCCCTCGATGAGGTC  
AATGGTGTGCTTTACCGCATGCGAAACGGCAAGATCGATGGTCGTGTGGCGATTTCGTTTC

40 (I) Angaben zur SEQ ID NR. 4:

(1) Sequenzcharakteristika:

(A) Länge: 1002 Basenpaare  
45 (B) Typ: Nucleinsäure  
(C) Strangtyp: Doppelstrang

12

(D) Topologie: linear

(2) Molekülart: DNA

(3) hypothetisch: nein

5 (4) Antisense: nein

(5) Herkunft:

(D) Organismus: *Corynebacterium glutamicum*

10

(6) Beschreibung der Sequenz: SEQ ID NR. 4:

AAGTGGAGCTCGCGCGCCTGCAGGTGACACTAGTGGATCAGAGGCATACTCCGGCGGACTCACC  
TACTCCGGACACCCACTTGCAGTAGCACCCGCCAAGGCAGCGCTGGAGATTTACGCGGAAGGAGA  
15 GATCATTCCACGCGTAGCTCGACTTGGCGCTGAACTGATCGAACCTCGCCTTCGTGAACTAGCGG  
AAGAAAACGTAGCGATCGCTGACGTGCGGGGCATCGGATTCTTCTGGGCAGTGGAGTTCAATGCA  
GACGCCACTGCCATGGCTGCCGGTGCTGCAGAATTCAAGGAACGCGGCGTGTGGCCGATGATCTC  
CGGCAACCGATTCCACATCGCGCCGCGCTGACCACCACTGATGACGAATTGGTAGCACTGCTGG  
ACGCGGTGGAAGCTGCAGCCCAAGCTGTGAGCTGACCTTCGCTGGGGCGTTGTTCTAAGTTTTC  
20 TAGATAACAAGGCCAGCACAGACCACCATNTCTACGACCCCAAAAACCGACTCCAAGCTCCGCGG  
CGACNAANCCGCGCTCGCGCCACCGACCAAGCAGCCGGTCCAGGTTTAAAGATTTTGCTTTTCGA  
CGCTCCCCTCCACCTCATTCAATGCGGCGGAAGGGATTTCTTGCATGTTTAAGCCTATAGGAAA  
AAGTGTTTGCATATCACCTTGTATTCCAACACTTGAGCGGGTAGANTGGGTGGTAACNACCCNG  
GGAAAGGGGGAAGACACCATGAGCATCNCCACNCACNTCCAAGCNCTCNCCACAGCANTCAACGC  
25 CATCNACAACCATTTGGNCAGCATGCTCNAACATNGTGTTCNCCANACAATANANGGCNTNNA  
NCCCGACTCANCNCCTANAAANACNCCTTCACCACACGCCNCCTTCGNCCCCAAACCAAACCCTCG  
CCNAAGCNCAACNCGCCACNCATTNGCTCCCCNCCTCNTNNATACCTNCCNCCCTTCGGATATCN  
AGCANGCGCCNCACCGNTCATTTNCCN

30 (I) Angaben zur SEQ ID NR. 5:

(1) Sequenzcharakteristika:

(A) Länge: 1007 Basenpaare

35 (B) Typ: Nucleinsäure

(C) Strangtyp: Doppelstrang

(D) Topologie: linear

(2) Molekülart: DNA

40 (3) hypothetisch: nein

(4) Antisense: nein

(5) Herkunft:

45 (E) Organismus: *Corynebacterium glutamicum*

## 13

(6) Beschreibung der Sequenz: SEQ ID NR. 5:

TCCCNATTGGGTACCTTACCTGGTACCCACCCGGGTGGAAAATCGATGGGCCCCGCGGCCGCTCTA  
GAAGTACTCTCGAGAAGCTTTTTGAATTCTTTGGATCCGAGCTGAACACATGGGTGATGTTTTTT  
5 TGAACCAGCACTGAGGCTGCGCTGGCCGCCTGTTGAAAGCCCAGGTCAGACAGCTCTGTGTCCAA  
TTGTCCCTGCATTCGGGACGTGGCGTTGTATTCAGTCTGCCCCGTGTCGGAGCAGAATCAGGCGAC  
GAGTCACAGGACTACCTCTTAATCGTCCTCGTAGCCAGGCTCGTATTCAGCTGGCAAAGGTGGGA  
GTTTCATCAATGCTGTCGATGTTGCGGATATCCGCCTCATCAGACCAGGAGGATNCACGCNTGAAG  
GTTTCAAGTCCTTCAATTTCAATGAGTGGGCAGTCNCGGTACAGACNATCCANTCCGTATAACTC  
10 GCGCTCTGCCTGTGCTGAACGTGGATAACAACCNATCCGTAGTCAAGGAGAACCCAACNGTTTT  
CGCGGTTGCCTTCACGGCGCTTAGGCTCGAAACCAGCCTTGGTCATCTNCATCTTCGATCTCCTC  
NACAATGGCGCCACCTTGGCGCTCATTTGTCCGCAGATGCAACNACNAANCAATTCTCGTGATT  
GNCGATCACTGTCNNAAACATCCAATNACAGCGATGTCNNCNGCCTTTCTTTTCNTGCCGCTGCT  
TTCGCCNCCATGGTCCCGAAGCCGATCGANTCCTCCATNTGCANATCAAAATTCNNTAAANCAGC  
15 TNCNTGTNGTTCNCNACCCNCTTTTTANGGTCCGAAACCNACCCCTNCNGAAANAATCCCCACGTC  
AACCTTCCCTNTTTCCCNCTANACCGGGTGATTCNCCTACTTTNGGNTCGAATTTAAACTTTTNA  
NCANATTTCTCTNGTTTTGGGCCTTGGGATCATTCCTTATTCGATCCTNCTGGTCAAAAATTG  
GGNTTNGGCTATTCTTCNCCACCCCCCANGGA

20 (I) Angaben zur SEQ ID NR. 6:

(1) Sequenzcharakteristika:

(A) Länge: 748 Basenpaare  
25 (B) Typ: Nucleinsäure  
(C) Strangtyp: Doppelstrang  
(D) Topologie: linear

(2) Molekülart: DNA  
30 (3) hypothetisch: nein  
(4) Antisense: nein

(5) Herkunft:

35 (F) Organismus: *Corynebacterium glutamicum*

(6) Beschreibung der Sequenz: SEQ ID NR. 6:

TTGANNCNTTNNNGGAGCTCCCCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCGACACGCTGAC  
40 ATCACCAGGCTGCAAATCAAGGCCAAGCGCAGCCGAGCATTTATCTCCCGTGCCTGCAGCAACTT  
TCACTCCACCTGGAGTTGTTCCCGCAATCGCATTTGGGGCCAGGAGTTCAGGAAGTTCCACCTCA  
TGGCCCAGCGCCAAGGCAGCTAGATCGGTGCGCCACGCACGATCACGCGTGCTGTAGTAGCCCGT  
TCCAGAAGCATCACCATGGTCGGTGACTTTGCGTCCGCGTCCCATCAAATGCCAGGTGAGGAAAT  
CATGAGGCAACATCACCGACGCCGTGCGCGCTGCATTTTCTGGTTCATGATCACGCATCCACCGC  
45 ATTTTGGTGGCAGTTAAAGAAGCAACATACAACTTCCCGTGGCATCTACCGCAGCCTGATCGCC  
GCCGATCTCCTCATTTGAGATCCAACGCAGCCTGGGCAGAACGAGTGTCATTCCATAACAACGCCG  
GGCGAACGATTTTCATCGTTTTTCATCCAACGCCACCATGCCGTGCTGCTGGCCTGCAATAGATACA



## 14

GCGTCCGCGCGTTCTAACAACCCCTCGGTAGCTTGATCCAGCGCAGCGATCCACGCACGTGGATC  
TACTTCGACCCGTTCTGGGGTGACTCGCGCGGCTTCGTCGATACCTGGCCGGTGGCGGCGTCACAA  
GCAAAAGCCTTGCAGGAATGGGTGGGAACATC

## 5 (I) Angaben zur SEQ ID NR. 7:

## (1) Sequenzcharakteristika:

- (A) Länge: 648 Basenpaare  
10 (B) Typ: Nucleinsäure  
(C) Strangtyp: Doppelstrang  
(D) Topologie: linear  
(2) Molekülart: DNA  
(3) hypothetisch: nein  
15 (4) Antisense: nein  
  
(5) Herkunft:  
(G) Organismus: *Corynebacterium glutamicum*

## 20 (6) Beschreibung der Sequenz: SEQ ID NR. 7:

TGCAGCCCGGGGGATCACCGACGCCAAGGCTACGTAGGAATCCCCTTCCCCGACACCATCGTGCG  
CATCGCAAACCCAGAAAACCTCGACGAAACCATGCCCCGACGGCAGCGAAGGCGAAGTCCTAGTCA  
AGGGCCACAGGTGTTCAAGGGTTACCTCAACCAGGAAGAAGCCACCAAGAACAGCTTCCACGGC  
25 GAGTGGTACCGCACCGGCGACGTCTGGAGTGATGGAAGAAGACGGGTTCATCCGCCTAGTTGCTCG  
CATCAAGGAAGTCATCATCACTGGCGGTTTCAACGTGTACCCAGCTGAGGTTGAAGAAGTCCTCG  
CAGAGCACCCAGACATTGAAGATTCCGCGAGTCGTTGGTATCCCGCGTGAAGACGGCTCCGAAAAC  
GTCGTTGCTGCATCACTTTGGTGGAAGGTGCAGCGCTGGATCCGGATGGCCTGAAGGAATTGCC  
GCAAGAACCTACCCGCTCAAGGTTCCGCGCACTTTCTACCACTTTGAGGAGATGCCGCGGGATCA  
30 GATGGCAAGATTAGGCGTCGTGAAGTGCANGCGGAGTTGTTGAAGAACTCGGCAGTNACGCCGAT  
TAAGAGGTCAGTTTCCAAATGGCACTTACCAATTGGNCTAGTTACCCCCANAAGCATTTTGAGGG  
TTCCACTTTTACCCAGTGGGNTGTGTGATCCTNT

## (I) Angaben zur SEQ ID NR. 8:

35

## (1) Sequenzcharakteristika:

- (A) Länge: 698 Basenpaare  
(B) Typ: Nucleinsäure  
40 (C) Strangtyp: Doppelstrang  
(D) Topologie: linear  
  
(2) Molekülart: DNA  
(3) hypothetisch: nein  
45 (4) Antisense: nein

15

(5) Herkunft:

(H) Organismus:

*Corynebacterium glutamicum*

5 (6) Beschreibung der Sequenz: SEQ ID NR. 8:

GCAGCCCGGGGATCCTTGGTGNCACCACCCTGGACATGCTCAAGATGGAACAGCAAATCGACTC  
CCTGGCACCAGGCGATGCGAAGCGCTACATGCACCACTACAACCTCCCTCCATACTCCACCGGTG  
AAACCGGTCTGTGGGCTCACCAAAGCGCCGCGAAATCGGCCACGGTGCACCTTGCAGAACGCGCA  
10 GTTTTGCCAGTAATCCCATCCCGTGAGGAATTCCCATACGCAATCCGTCAGGTCTCTGAAGCTCT  
GGGCTCCAACGGCTCCACCTCCATGGGCTCTGTCTGTGCATCCACTCTGTCCCTGTACAACGCTG  
GTGTTCCACTGAAGGCACCTGTTGCAGGTATCGCCATGGGACTTGTTTCCGGTGAAATCGACGGC  
AAGACCGAGTACGTTGCACTGACCGACATCCTCGGCGCAGAAGACGCATTTCGGCGACATGGACTT  
CAAGGTTGCCGGCACCGCAGACTTCATACCGNACTTCAGCTGGACACCAAGCTGGACNGCATTCC  
15 TTCAAGGTGCTCTCCGATGCGCTTGAGCANGCACGCGATNCCGACTGACATCTGAACACATGGCT  
GATGTATCAACGGACCTTGATGAGATGAGCAAGTTCGTTCTGCATAACCACCGNGAAATCCCATGG  
CAAAATCGNGACTGTGACCAAGGGTAGACATTACGCTTTACNATTTCG

(I) Angaben zur SEQ ID NR. 9:

20

(1) Sequenzcharakteristika:

(A) Länge: 1159 Basenpaare  
(B) Typ: Nucleinsäure  
25 (C) Strangtyp: Doppelstrang  
(D) Topologie: linear

(2) Molekülart: DNA  
(3) hypothetisch: nein  
30 (4) Antisense: nein

(5) Herkunft:

(I) Organismus:

*Corynebacterium glutamicum*

35

(6) Beschreibung der Sequenz: SEQ ID NR. 9:

TTNANNCGTTTGGAGCTCCCCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCACACAAAATGATT  
AGATTGTGTGCGAATTCATCCAATTGCTGTCTATATGCAGTACGCATGGCAACACTATAAGGCGA  
40 TAATGGTATTTCTGCAGGCCTAAAACACCCCTTTAAGATTGAATCACCTAATAATGGGGGATAGC  
CAACTATTGGCGGGGGTAAGT  
ATTAAATTAACTACCCTCCGGAGTTTTCCATTTCTGCGCCTTTAGGAACAGTGGCATCCTCAGGA  
TCGTCCAACCAGCCATCTGGAAGTGCCACCTTTGCAGGAGCGCCCTGTCGACCTCGTGCACCTTC  
CGCGTCCTCTGCCTTGGCGGTGRAGTCAGCCCATGGTGCTAGGAGATCCTCAAGCTCCACATAGG  
45 TGGAACCTTGGCCAGATTGGAGCGGAATTCGCCGCCAACAGGGAAACCGCGCAGGTCCAACCCA  
TGTGCTTACGCAGATCGCGCAGCCCCCTTGGTTTCGCCATCATGCTGCATGAGGAGTTCTGCGTGG  
CGCAGGATGATTTGGGTAACTTCGCCGAAGGTAGGCTCCTCTGGGATTTCTTCTCCACGAACAGC

16

AGCAGCAGCTCAGCAAAGAGCCAAGGCCTGCCCAGGCAACCACgGCCCAACCACGACGCCATCGC  
AGCCAGTTTGCTCCATCATGCGCGTTGCATCGGATGCCGCGAAAATATCGCCATTGCCCAAACCT  
GGGATGCCGGTATCTGCCAAATGCTCYTTCAGGCGCGCGATCTYGTTCCAATCAGCCTCACCGGA  
ATAGCGCTGCGCCGCGAGTGCGGGCGTGAAGCGCTACGGACTTCGCGCCGGCGTCGACAGCAATGC  
5 GTCCAGCATCCAAGTGAGTATGGTGCTCATCATCAATACCAACGCGGAACTTCACCGTCACCGGA  
ATGTCCGTGCCTTCCGTAGCCTTCACAGCCGCGGAAACGATGTTTTCAAACAAACGGCGCTTGTA  
AGGAATCGCAGAACCGCCACCCCGGCGCGTGACCTTTGGAACCGGGCAGCCAAAGTTCATATCAA  
TATGATCCCCCGGGCTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGG  
CCCCGTACCCAGCTTGTGTTCCAANGNTCCAA

10

(I) Angaben zur SEQ ID NR. 10:

(1) Sequenzcharakteristika:

15 (A) Länge: 761 Basenpaare  
(B) Typ: Nucleinsäure  
(C) Strangtyp: Doppelstrang  
(D) Topologie: linear

20 (2) Molekülart: DNA  
(3) hypothetisch: nein  
(4) Antisense: nein

(5) Herkunft:

25

(J) Organismus: *Corynebacterium glutamicum*

(6) Beschreibung der Sequenz: SEQ ID NR. 10:

30 TTGAANCCTTANNGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCTCGATTCACT  
CGAGCTTGATGAAACTGTCAAGGTCGTTGCCTTCACTCACCAGTCCAATGTGACCGGTGCTGTGG  
CTGATGTTCCAGAGTTGGTTCGTCGTGCCAAGGCTGTCGGCGCTCTCACGGTGCTTGATGCGTGC  
CAGTCTGTTCCCTCATATGCCAGTGAATTTCCACGAGCTGGATGTAGATTTCTCTGCATTCTCTGG  
CCATAAGATGCTGGGACCTGCAGGCGTGGGCGTTGTGTATGCAAAGTCCCCAATCTTGGATGAAC  
35 TGCCACCATTTTTGACTGGTGGTTCCATGATTGAAGTTGTCACCATGGAGGGTTCCACCTACGCT  
GCCGCACCTCAACGTTTTGAGGCCGGCACGCAGATGACCAGCCAGGTTGTGGGCTTGGGTGCTGC  
CGTGGACATGCTGAATGAAATCGGTATGGAAGCAATCGCAGCNGCATGAGCACGCATTGACTGCT  
TACGCGTTGGAAAAGCTCACGGCAATTAAAGGGACTAACCATTGCTGGTCCTTTTACTGTCAGAG  
CATCGCGNGGTGCAATCAGCTTCNGTGTCNANGGCATTCACCNACACGATCTANGGCAAAGTGC  
40 TTGACCATCAGGGCGTGAATATTCGNGTCGGGCACCACTGTGCGTGGGCCTGCACCGCANCATT  
GAACGTNCAATNGNANACAAGAGCATTTTTCTATCTCTATTACACC

(I) Angaben zur SEQ ID NR. 11:

45 (1) Sequenzcharakteristika:

(A) Länge: 791 Basenpaare

17

(B) Typ: Nucleinsäure  
(C) Strangtyp: Doppelstrang  
(D) Topologie: linear

5 (2) Molekülart: DNA  
(3) hypothetisch: nein  
(4) Antisense: nein

(5) Herkunft:

10

(K) Organismus: *Corynebacterium glutamicum*

(6) Beschreibung der Sequenz: SEQ ID NR. 11:

15 TTGACCCTTTAGCTGGGTACCGGGCCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGA  
ATTCTGTCAGCCCGGGGGATCTTCATCGCCAACAACTGACCGCGGGAAACGATCATCTTCTCAA  
ATTCTGTGAGCTTTTCCAGCGCCTTGTCGACGGGTGGCCCCACGATCTCCTCGGCCATAACGGAC  
GTGGAGGCCTGGCTGATTGAGCAACCAACTGCTTCGTAGGAGACGTCCTCCACGGTGGAGCCGTC  
CTCAGACAGCTTCACGCGCAGAGTCAATTCGTCGCCACAAGAAGGGTTGACGTGGTGAACCTCAG  
20 CATCGAAAGGATCCCGAAGGCCCTTGCTGCTGTGGGTTTTTGTAGTGGTCCAGGATCACCTCCTGG  
TACATCTGCTCAAGGTTCACTCAACTCCAAAGAATTGCTTGGCCTTCTCGATCGCTGCCGCG  
AGGCGGTGATTTCTTCGAAGGTGTTATAGAGATAGAAAGATGCTCTTGCTGTCGATTGTACCGT  
TCATGCTGCGGTGCACGGCCACGCGCAGTGGTGGNCGACGCCGGATATTCACGCCCTGATCGTCA  
AGCACTTGGCCTAANC GTGTGGGTGAATGCCTCGACACCGAACTGATGCACCGGCGNCTGCTNTN  
25 CATCAAAGGACCANCNATGGGTAAAGTCCTTAATGCCGNGAGCTTTTCAACGCGTAAGCAGGTAA  
TGCNNCTATGCNCTGCGATGNTTTCATACCGATTNNTTAAGANTNTCCCCGGNGTNCNCCNANCCC  
NAACTGGTTN

30

35

40

45

## Patentansprüche

1. Ein gereinigtes Polynucleotid mit einer Nukleinsäuresequenz,  
5 die aus der folgenden Gruppe ausgewählt ist: SEQ ID NR. 1, SEQ ID NR. 2, SEQ ID NR. 3, SEQ ID NR. 4; SEQ ID NR. 5, SEQ ID NR. 6, SEQ ID NR. 7, SEQ ID NR. 8, SEQ ID NR. 9, SEQ ID NR. 10, SEQ ID NR. 11.
- 10 2. Ein Expressions-Vektor mit einem dem Anspruch 1 entsprechenden Polynucleotid.
3. Eine Wirtszelle, die mit dem Expressions-Vektor aus Anspruch 2 transformiert ist.
- 15 4. Eine Methode zur Herstellung und Reinigung eines Polypeptids, die aus folgenden Schritten besteht:
- (a) Kultivierung der Wirtszelle aus Anspruch 3 unter Bedin-  
20 gungen, die für die Expression des Peptids geeignet sind; und
- (b) Gewinnung des Polypeptids aus der Wirtszellkultur.

25

30

35

40

45

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
4 January 2001 (04.01.2001)

PCT

(10) International Publication Number  
**WO 01/00843 A2**

(51) International Patent Classification<sup>7</sup>: **C12N 15/31**,  
15/61, 1/21, 9/90, C07K 14/34, C12P 13/08, C12Q 1/68

(21) International Application Number: **PCT/IB00/00923**

(22) International Filing Date: **23 June 2000 (23.06.2000)**

(25) Filing Language: **English**

(26) Publication Language: **English**

(30) Priority Data:

60/141,031	25 June 1999 (25.06.1999)	US
199 30 476.9	1 July 1999 (01.07.1999)	DE
60/142,101	2 July 1999 (02.07.1999)	US
199 31 415.2	8 July 1999 (08.07.1999)	DE
199 31 418.7	8 July 1999 (08.07.1999)	DE
199 31 419.5	8 July 1999 (08.07.1999)	DE
199 31 420.9	8 July 1999 (08.07.1999)	DE
199 31 424.1	8 July 1999 (08.07.1999)	DE
199 31 428.4	8 July 1999 (08.07.1999)	DE
199 31 434.9	8 July 1999 (08.07.1999)	DE
199 31 435.7	8 July 1999 (08.07.1999)	DE
199 31 443.8	8 July 1999 (08.07.1999)	DE
199 31 453.5	8 July 1999 (08.07.1999)	DE
199 31 457.8	8 July 1999 (08.07.1999)	DE
199 31 465.9	8 July 1999 (08.07.1999)	DE
199 31 478.0	8 July 1999 (08.07.1999)	DE
199 31 510.8	8 July 1999 (08.07.1999)	DE
199 31 541.8	8 July 1999 (08.07.1999)	DE
199 31 573.6	8 July 1999 (08.07.1999)	DE
199 31 592.2	8 July 1999 (08.07.1999)	DE
199 31 632.5	8 July 1999 (08.07.1999)	DE
199 31 634.1	8 July 1999 (08.07.1999)	DE
199 31 636.8	8 July 1999 (08.07.1999)	DE
199 32 125.6	9 July 1999 (09.07.1999)	DE
199 32 126.4	9 July 1999 (09.07.1999)	DE
199 32 130.2	9 July 1999 (09.07.1999)	DE
199 32 186.8	9 July 1999 (09.07.1999)	DE
199 32 206.6	9 July 1999 (09.07.1999)	DE
199 32 227.9	9 July 1999 (09.07.1999)	DE
199 32 228.7	9 July 1999 (09.07.1999)	DE
199 32 229.5	9 July 1999 (09.07.1999)	DE
199 32 230.9	9 July 1999 (09.07.1999)	DE
199 32 922.2	14 July 1999 (14.07.1999)	DE
199 32 926.5	14 July 1999 (14.07.1999)	DE
199 32 928.1	14 July 1999 (14.07.1999)	DE

199 33 004.2	14 July 1999 (14.07.1999)	DE
199 33 005.0	14 July 1999 (14.07.1999)	DE
199 33 006.9	14 July 1999 (14.07.1999)	DE
60/148,613	12 August 1999 (12.08.1999)	US
199 40 764.9	27 August 1999 (27.08.1999)	DE
199 40 765.7	27 August 1999 (27.08.1999)	DE
199 40 766.5	27 August 1999 (27.08.1999)	DE
199 40 832.7	27 August 1999 (27.08.1999)	DE
199 41 378.9	31 August 1999 (31.08.1999)	DE
199 41 379.7	31 August 1999 (31.08.1999)	DE
199 41 394.0	31 August 1999 (31.08.1999)	DE
199 41 396.7	31 August 1999 (31.08.1999)	DE
199 41 380.0	31 August 1999 (31.08.1999)	DE
199 42 077.7	3 September 1999 (03.09.1999)	DE
199 42 129.3	3 September 1999 (03.09.1999)	DE
199 42 076.9	3 September 1999 (03.09.1999)	DE
199 42 079.3	3 September 1999 (03.09.1999)	DE
199 42 086.6	3 September 1999 (03.09.1999)	DE
199 42 087.4	3 September 1999 (03.09.1999)	DE
199 42 088.2	3 September 1999 (03.09.1999)	DE
199 42 095.5	3 September 1999 (03.09.1999)	DE
199 42 124.2	3 September 1999 (03.09.1999)	DE
60/187,970	9 March 2000 (09.03.2000)	US

(71) Applicant: **BASF AKTIENGESSELLSCHAFT**  
[DE/DE]; D-67056 Ludwigshafen (DE).

(72) Inventors: **POMPEJUS**, Markus; Wenjenstr. 21,  
D-67251 Freinsheim (DE). **KRÖGER**, Burkhard; Im  
Waldhof 1, D-67117 Limburgerhof (DE). **SCHRÖDER**,  
Hartwig; Goethestr. 5, D-69226 Nussloch (DE).  
**ZELDER**, Oskar; Rossmarktstr. 27, D-67346 Speyer  
(DE). **HABERHAUER**, Gregor; Moselstr. 42, D-67117  
Limburgerhof (DE).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,  
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,  
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,  
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,  
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,  
TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,

[Continued on next page]

(54) Title: **CORYNEBACTERIUM GLUTAMICUM GENES ENCODING METABOLIC PATHWAY PROTEINS**

(57) Abstract: Isolated nucleic acid molecules, designated MP nucleic acid molecules, which encode novel MP proteins from *Corynebacterium glutamicum* are described. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing MP nucleic acid molecules, and host cells into which the expression vectors have been introduced. The invention still further provides isolated MP proteins, mutated MP proteins, fusion proteins, antigenic peptides and methods for the improvement of production of a desired compound from *C. glutamicum* based on genetic engineering of MP genes in this organism.



WO 01/00843 A2





IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

**Published:**

- *Without international search report and to be republished upon receipt of that report.*

***CORYNEBACTERIUM GLUTAMICUM* GENES ENCODING METABOLIC  
PATHWAY PROTEINS**

**Related Applications**

The present application claims priority to prior filed U.S. Provisional Patent  
5 Application Serial No. 60/141031, filed June 25 , 1999, U.S. Provisional Patent  
Application Serial No. 60/142101, filed July 2, 1999, U.S. Provisional Patent  
Application Serial No. 60/148613, filed August 12, 1999, and also to U.S. Provisional  
Patent Application Serial No. 60/187970, filed March 9, 2000. The present application  
also claims priority to prior filed German Patent Application No. 19930476.9, filed July  
10 1, 1999, German Patent Application No. 19931415.2, filed July 8, 1999, German Patent  
Application No. 19931418.7, filed July 8, 1999, German Patent Application No.  
19931419.5, filed July 8, 1999, German Patent Application No. 19931420.9, filed July  
8, 1999, German Patent Application No. 19931424.1, filed July 8, 1999, German Patent  
Application No. 19931428.4, filed July 8, 1999, German Patent Application No.  
15 19931434.9, filed July 8, 1999, German Patent Application No. 19931435.7, filed July  
8, 1999, German Patent Application No. 19931443.8, filed July 8, 1999, German Patent  
Application No. 19931453.5, filed July 8, 1999, German Patent Application No.  
19931457.8, filed July 8, 1999, German Patent Application No. 19931465.9, filed July  
8, 1999, German Patent Application No. 19931478.0, filed July 8, 1999, German Patent  
20 Application No. 19931510.8, filed July 8, 1999, German Patent Application No.  
19931541.8, filed July 8, 1999, German Patent Application No. 19931573.6, filed July  
8, 1999, German Patent Application No. 19931592.2, filed July 8, 1999, German Patent  
Application No. 19931632.5, filed July 8, 1999, German Patent Application No.  
19931634.1, filed July 8, 1999, German Patent Application No. 19931636.8, filed July  
25 8, 1999, German Patent Application No. 19932125.6, filed July 9, 1999, German Patent  
Application No. 19932126.4, filed July 9, 1999, German Patent Application No.  
19932130.2, filed July 9, 1999, German Patent Application No. 19932186.8, filed July  
9, 1999, German Patent Application No. 19932206.6, filed July 9, 1999, German Patent  
Application No. 19932227.9, filed July 9, 1999, German Patent Application No.  
30 19932228.7, filed July 9, 1999, German Patent Application No. 19932229.5, filed July  
9, 1999, German Patent Application No. 19932230.9, filed July 9, 1999, German Patent  
Application No. 19932922.2, filed July 14, 1999, German Patent Application No.

19932926.5, filed July 14, 1999, German Patent Application No. 19932928.1, filed July 14, 1999, German Patent Application No. 19933004.2, filed July 14, 1999, German Patent Application No. 19933005.0, filed July 14, 1999, German Patent Application No. 19933006.9, filed July 14, 1999, German Patent Application No. 19940764.9, filed  
5 August 27, 1999, German Patent Application No. 19940765.7, filed August 27, 1999, German Patent Application No. 19940766.5, filed August 27, 1999, German Patent Application No. 19940832.7, filed August 27, 1999, German Patent Application No. 19941378.9, filed August 31, 1999, German Patent Application No. 19941379.7, filed August 31, 1999, German Patent Application No. 19941380.0, filed August 31, 1999,  
10 German Patent Application No. 19941394.0, filed August 31, 1999, German Patent Application No. 19941396.7, filed August 31, 1999, German Patent Application No. 19942076.9, filed September 3, 1999, German Patent Application No. 19942077.7, filed September 3, 1999, German Patent Application No. 19942079.3, filed September 3, 1999, German Patent Application No. 19942086.6, filed September 3, 1999, German  
15 Patent Application No. 19942087.4, filed September 3, 1999, German Patent Application No. 19942088.2, filed September 3, 1999, German Patent Application No. 19942095.5, filed September 3, 1999, German Patent Application No. 19942124.2, filed September 3, 1999, and German Patent Application No. 19942129.3, filed September 3, 1999. The entire contents of all of the aforementioned applications are hereby expressly  
20 incorporated herein by this reference.

### **Background of the Invention**

Certain products and by-products of naturally-occurring metabolic processes in cells have utility in a wide array of industries, including the food, feed, cosmetics, and  
25 pharmaceutical industries. These molecules, collectively termed 'fine chemicals', include organic acids, both proteinogenic and non-proteinogenic amino acids, nucleotides and nucleosides, lipids and fatty acids, diols, carbohydrates, aromatic compounds, vitamins and cofactors, and enzymes. Their production is most conveniently performed through large-scale culture of bacteria developed to produce  
30 and secrete large quantities of a particular desired molecule. One particularly useful organism for this purpose is *Corynebacterium glutamicum*, a gram positive, nonpathogenic bacterium. Through strain selection, a number of mutant strains have

been developed which produce an array of desirable compounds. However, selection of strains improved for the production of a particular molecule is a time-consuming and difficult process.

## 5 Summary of the Invention

The invention provides novel bacterial nucleic acid molecules which have a variety of uses. These uses include the identification of microorganisms which can be used to produce fine chemicals, the modulation of fine chemical production in *C. glutamicum* or related bacteria, the typing or identification of *C. glutamicum* or related  
10 bacteria, as reference points for mapping the *C. glutamicum* genome, and as markers for transformation. These novel nucleic acid molecules encode proteins, referred to herein as metabolic pathway (MP) proteins.

*C. glutamicum* is a gram positive, aerobic bacterium which is commonly used in industry for the large-scale production of a variety of fine chemicals, and also for the  
15 degradation of hydrocarbons (such as in petroleum spills) and for the oxidation of terpenoids. The MP nucleic acid molecules of the invention, therefore, can be used to identify microorganisms which can be used to produce fine chemicals, e.g., by fermentation processes. Modulation of the expression of the MP nucleic acids of the invention, or modification of the sequence of the MP nucleic acid molecules of the  
20 invention, can be used to modulate the production of one or more fine chemicals from a microorganism (e.g., to improve the yield or production of one or more fine chemicals from a *Corynebacterium* or *Brevibacterium* species).

The MP nucleic acids of the invention may also be used to identify an organism as being *Corynebacterium glutamicum* or a close relative thereof, or to identify the  
25 presence of *C. glutamicum* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *C. glutamicum* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *C. glutamicum* gene which is unique to this organism, one can ascertain  
30 whether this organism is present. Although *Corynebacterium glutamicum* itself is nonpathogenic, it is related to species pathogenic in humans, such as *Corynebacterium*

*diphtheriae* (the causative agent of diphtheria); the detection of such organisms is of significant clinical relevance.

The MP nucleic acid molecules of the invention may also serve as reference points for mapping of the *C. glutamicum* genome, or of genomes of related organisms.

5 Similarly, these molecules, or variants or portions thereof, may serve as markers for genetically engineered *Corynebacterium* or *Brevibacterium* species.

The MP proteins encoded by the novel nucleic acid molecules of the invention are capable of, for example, performing an enzymatic step involved in the metabolism of certain fine chemicals, including amino acids, vitamins, cofactors, nutraceuticals,

10 nucleotides, nucleosides, and trehalose. Given the availability of cloning vectors for use in *Corynebacterium glutamicum*, such as those disclosed in Sinskey *et al.*, U.S. Patent No. 4,649,119, and techniques for genetic manipulation of *C. glutamicum* and the related *Brevibacterium* species (*e.g.*, *lactofermentum*) (Yoshihama *et al.*, *J. Bacteriol.* 162: 591-597 (1985); Katsumata *et al.*, *J. Bacteriol.* 159: 306-311 (1984); and  
15 Santamaria *et al.*, *J. Gen. Microbiol.* 130: 2237-2246 (1984)), the nucleic acid molecules of the invention may be utilized in the genetic engineering of this organism to make it a better or more efficient producer of one or more fine chemicals.

This improved production or efficiency of production of a fine chemical may be due to a direct effect of manipulation of a gene of the invention, or it may be due to an  
20 indirect effect of such manipulation. Specifically, alterations in *C. glutamicum* metabolic pathways for amino acids, vitamins, cofactors, nucleotides, and trehalose may have a direct impact on the overall production of one or more of these desired compounds from this organism. For example, optimizing the activity of a lysine biosynthetic pathway protein or decreasing the activity of a lysine degradative pathway  
25 protein may result in an increase in the yield or efficiency of production of lysine from such an engineered organism. Alterations in the proteins involved in these metabolic pathways may also have an indirect impact on the production or efficiency of production of a desired fine chemical. For example, a reaction which is in competition for an intermediate necessary for the production of a desired molecule may be eliminated, or a  
30 pathway necessary for the production of a particular intermediate for a desired compound may be optimized. Further, modulations in the biosynthesis or degradation of, for example, an amino acid, a vitamin, or a nucleotide may increase the overall

ability of the microorganism to rapidly grow and divide, thus increasing the number and/or production capacities of the microorganism in culture and thereby increasing the possible yield of the desired fine chemical.

The nucleic acid and protein molecules of the invention may be utilized to  
5 directly improve the production or efficiency of production of one or more desired fine chemicals from *Corynebacterium glutamicum*. Using recombinant genetic techniques well known in the art, one or more of the biosynthetic or degradative enzymes of the invention for amino acids, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, or trehalose may be manipulated such that its function is modulated. For example, a  
10 biosynthetic enzyme may be improved in efficiency, or its allosteric control region destroyed such that feedback inhibition of production of the compound is prevented. Similarly, a degradative enzyme may be deleted or modified by substitution, deletion, or addition such that its degradative activity is lessened for the desired compound without impairing the viability of the cell. In each case, the overall yield or rate of production of  
15 the desired fine chemical may be increased.

It is also possible that such alterations in the protein and nucleotide molecules of the invention may improve the production of other fine chemicals besides the amino acids, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, and trehalose through indirect mechanisms. Metabolism of any one compound is necessarily  
20 intertwined with other biosynthetic and degradative pathways within the cell, and necessary cofactors, intermediates, or substrates in one pathway are likely supplied or limited by another such pathway. Therefore, by modulating the activity of one or more of the proteins of the invention, the production or efficiency of activity of another fine chemical biosynthetic or degradative pathway may be impacted. For example, amino  
25 acids serve as the structural units of all proteins, yet may be present intracellularly in levels which are limiting for protein synthesis; therefore, by increasing the efficiency of production or the yields of one or more amino acids within the cell, proteins, such as biosynthetic or degradative proteins, may be more readily synthesized. Likewise, an alteration in a metabolic pathway enzyme such that a particular side reaction becomes  
30 more or less favored may result in the over- or under-production of one or more compounds which are utilized as intermediates or substrates for the production of a desired fine chemical.



This invention provides novel nucleic acid molecules which encode proteins, referred to herein as metabolic pathway proteins (MP), which are capable of, for example, performing an enzymatic step involved in the metabolism of molecules important for the normal functioning of cells, such as amino acids, vitamins, cofactors, nucleotides and nucleosides, or trehalose. Nucleic acid molecules encoding an MP protein are referred to herein as MP nucleic acid molecules. In a preferred embodiment, the MP protein performs an enzymatic step related to the metabolism of one or more of the following: amino acids, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, and trehalose. Examples of such proteins include those encoded by the genes set forth in Table 1.

Accordingly, one aspect of the invention pertains to isolated nucleic acid molecules (*e.g.*, cDNAs, DNAs, or RNAs) comprising a nucleotide sequence encoding an MP protein or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection or amplification of MP-encoding nucleic acid (*e.g.*, DNA or mRNA). In particularly preferred embodiments, the isolated nucleic acid molecule comprises one of the nucleotide sequences set forth as the odd-numbered SEQ ID NOs in the Sequence Listing (*e.g.*, SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7....), or the coding region or a complement thereof of one of these nucleotide sequences. In other particularly preferred embodiments, the isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes to or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80% or 90%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence set forth as an odd-numbered SEQ ID NO in the Sequence Listing (*e.g.*, SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7....), or a portion thereof. In other preferred embodiments, the isolated nucleic acid molecule encodes one of the amino acid sequences set forth as an even-numbered SEQ ID NO in the Sequence Listing (*e.g.*, SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8....). The preferred MP proteins of the present invention also preferably possess at least one of the MP activities described herein.

In another embodiment, the isolated nucleic acid molecule encodes a protein or portion thereof wherein the protein or portion thereof includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of the invention (*e.g.*, a

sequence having an even-numbered SEQ ID NO: in the Sequence Listing), *e.g.*, sufficiently homologous to an amino acid sequence of the invention such that the protein or portion thereof maintains an MP activity. Preferably, the protein or portion thereof encoded by the nucleic acid molecule maintains the ability to perform an enzymatic reaction in a amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway. In one embodiment, the protein encoded by the nucleic acid molecule is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90% and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an amino acid sequence of the invention (*e.g.*, an entire amino acid sequence selected from those having an even-numbered SEQ ID NO in the Sequence Listing). In another preferred embodiment, the protein is a full length *C. glutamicum* protein which is substantially homologous to an entire amino acid sequence of the invention (encoded by an open reading frame shown in the corresponding odd-numbered SEQ ID NOs in the Sequence Listing (*e.g.*, SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7....)).

In another preferred embodiment, the isolated nucleic acid molecule is derived from *C. glutamicum* and encodes a protein (*e.g.*, an MP fusion protein) which includes a biologically active domain which is at least about 50% or more homologous to one of the amino acid sequences of the invention (*e.g.*, a sequence of one of the even-numbered SEQ ID NOs in the Sequence Listing) and is able to catalyze a reaction in a metabolic pathway for an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose, or one or more of the activities set forth in Table 1, and which also includes heterologous nucleic acid sequences encoding a heterologous polypeptide or regulatory regions.

In another embodiment, the isolated nucleic acid molecule is at least 15 nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising a nucleotide sequence of the invention (*e.g.*, a sequence of an odd-numbered SEQ ID NO in the Sequence Listing). Preferably, the isolated nucleic acid molecule corresponds to a naturally-occurring nucleic acid molecule. More preferably, the isolated nucleic acid encodes a naturally-occurring *C. glutamicum* MP protein, or a biologically active portion thereof.

Another aspect of the invention pertains to vectors, *e.g.*, recombinant expression vectors, containing the nucleic acid molecules of the invention, and host cells into which such vectors have been introduced. In one embodiment, such a host cell is used to produce an MP protein by culturing the host cell in a suitable medium. The MP protein  
5 can be then isolated from the medium or the host cell.

Yet another aspect of the invention pertains to a genetically altered microorganism in which an MP gene has been introduced or altered. In one embodiment, the genome of the microorganism has been altered by introduction of a nucleic acid molecule of the invention encoding wild-type or mutated MP sequence as a  
10 transgene. In another embodiment, an endogenous MP gene within the genome of the microorganism has been altered, *e.g.*, functionally disrupted, by homologous recombination with an altered MP gene. In another embodiment, an endogenous or introduced MP gene in a microorganism has been altered by one or more point mutations, deletions, or inversions, but still encodes a functional MP protein. In still  
15 another embodiment, one or more of the regulatory regions (*e.g.*, a promoter, repressor, or inducer) of an MP gene in a microorganism has been altered (*e.g.*, by deletion, truncation, inversion, or point mutation) such that the expression of the MP gene is modulated. In a preferred embodiment, the microorganism belongs to the genus *Corynebacterium* or *Brevibacterium*, with *Corynebacterium glutamicum* being  
20 particularly preferred. In a preferred embodiment, the microorganism is also utilized for the production of a desired compound, such as an amino acid, with lysine being particularly preferred.

In another aspect, the invention provides a method of identifying the presence or activity of *Corynebacterium diphtheriae* in a subject. This method includes detection of  
25 one or more of the nucleic acid or amino acid sequences of the invention (*e.g.*, the sequences set forth in the Sequence Listing as SEQ ID NOs 1 through 1156) in a subject, thereby detecting the presence or activity of *Corynebacterium diphtheriae* in the subject.

Still another aspect of the invention pertains to an isolated MP protein or a  
30 portion, *e.g.*, a biologically active portion, thereof. In a preferred embodiment, the isolated MP protein or portion thereof can catalyze an enzymatic reaction involved in one or more pathways for the metabolism of an amino acid, a vitamin, a cofactor, a

nutraceutical, a nucleotide, a nucleoside, or trehalose. In another preferred embodiment, the isolated MP protein or portion thereof is sufficiently homologous to an amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: in the Sequence Listing) such that the protein or portion thereof maintains the ability to  
5 catalyze an enzymatic reaction involved in one or more pathways for the metabolism of an amino acid, a vitamin, a cofactor, a nutraceutical, a nucleotide, a nucleoside, or trehalose.

The invention also provides an isolated preparation of an MP protein. In preferred embodiments, the MP protein comprises an amino acid sequence of the  
10 invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing). In another preferred embodiment, the invention pertains to an isolated full length protein which is substantially homologous to an entire amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing) (encoded by an open reading frame set forth in a corresponding odd-numbered SEQ ID NO: of the  
15 Sequence Listing). In yet another embodiment, the protein is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90%, and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an entire amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing). In other embodiments, the isolated MP protein  
20 comprises an amino acid sequence which is at least about 50% or more homologous to one of the amino acid sequences of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing) and is able to catalyze an enzymatic reaction in an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway, or has one or more of the activities set forth in Table 1.

25 Alternatively, the isolated MP protein can comprise an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, *e.g.*, hybridizes under stringent conditions, or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80%, or 90%, and even more preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to a nucleotide sequence of one of  
30 the even-numbered SEQ ID NOs set forth in the Sequence Listing. It is also preferred that the preferred forms of MP proteins also have one or more of the MP bioactivities described herein.

The MP polypeptide, or a biologically active portion thereof, can be operatively linked to a non-MP polypeptide to form a fusion protein. In preferred embodiments, this fusion protein has an activity which differs from that of the MP protein alone. In other preferred embodiments, this fusion protein, when introduced into a *C. glutamicum* pathway for the metabolism of an amino acid, vitamin, cofactor, nutraceutical, results in increased yields and/or efficiency of production of a desired fine chemical from *C. glutamicum*. In particularly preferred embodiments, integration of this fusion protein into an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway of a host cell modulates production of a desired compound from the cell.

In another aspect, the invention provides methods for screening molecules which modulate the activity of an MP protein, either by interacting with the protein itself or a substrate or binding partner of the MP protein, or by modulating the transcription or translation of an MP nucleic acid molecule of the invention.

Another aspect of the invention pertains to a method for producing a fine chemical. This method involves the culturing of a cell containing a vector directing the expression of an MP nucleic acid molecule of the invention, such that a fine chemical is produced. In a preferred embodiment, this method further includes the step of obtaining a cell containing such a vector, in which a cell is transfected with a vector directing the expression of an MP nucleic acid. In another preferred embodiment, this method further includes the step of recovering the fine chemical from the culture. In a particularly preferred embodiment, the cell is from the genus *Corynebacterium* or *Brevibacterium*, or is selected from those strains set forth in Table 3.

Another aspect of the invention pertains to methods for modulating production of a molecule from a microorganism. Such methods include contacting the cell with an agent which modulates MP protein activity or MP nucleic acid expression such that a cell associated activity is altered relative to this same activity in the absence of the agent. In a preferred embodiment, the cell is modulated for one or more *C. glutamicum* amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathways, such that the yields or rate of production of a desired fine chemical by this microorganism is improved. The agent which modulates MP protein activity can be an agent which stimulates MP protein activity or MP nucleic acid expression.



Examples of agents which stimulate MP protein activity or MP nucleic acid expression include small molecules, active MP proteins, and nucleic acids encoding MP proteins that have been introduced into the cell. Examples of agents which inhibit MP activity or expression include small molecules, and antisense MP nucleic acid molecules.

5           Another aspect of the invention pertains to methods for modulating yields of a desired compound from a cell, involving the introduction of a wild-type or mutant MP gene into a cell, either maintained on a separate plasmid or integrated into the genome of the host cell. If integrated into the genome, such integration can be random, or it can take place by homologous recombination such that the native gene is replaced by the  
10 introduced copy, causing the production of the desired compound from the cell to be modulated. In a preferred embodiment, said yields are increased. In another preferred embodiment, said chemical is a fine chemical. In a particularly preferred embodiment, said fine chemical is an amino acid. In especially preferred embodiments, said amino acid is L-lysine.

15

#### **Detailed Description of the Invention**

The present invention provides MP nucleic acid and protein molecules which are involved in the metabolism of certain fine chemicals in *Corynebacterium glutamicum*, including amino acids, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, and  
20 trehalose. The molecules of the invention may be utilized in the modulation of production of fine chemicals from microorganisms, such as *C. glutamicum*, either directly (*e.g.*, where modulation of the activity of a lysine biosynthesis protein has a direct impact on the production or efficiency of production of lysine from that organism), or may have an indirect impact which nonetheless results in an increase of  
25 yield or efficiency of production of the desired compound (*e.g.*, where modulation of the activity of a nucleotide biosynthesis protein has an impact on the production of an organic acid or a fatty acid from the bacterium, perhaps due to improved growth or an increased supply of necessary co-factors, energy compounds, or precursor molecules). Aspects of the invention are further explicated below.

30



## I. Fine Chemicals

The term 'fine chemical' is art-recognized and includes molecules produced by an organism which have applications in various industries, such as, but not limited to, the pharmaceutical, agriculture, and cosmetics industries. Such compounds include organic acids, such as tartaric acid, itaconic acid, and diaminopimelic acid, both proteinogenic and non-proteinogenic amino acids, purine and pyrimidine bases, nucleosides, and nucleotides (as described *e.g.* in Kuninaka, A. (1996) Nucleotides and related compounds, p. 561-612, in Biotechnology vol. 6, Rehm *et al.*, eds. VCH: Weinheim, and references contained therein), lipids, both saturated and unsaturated fatty acids (*e.g.*, arachidonic acid), diols (*e.g.*, propane diol, and butane diol), carbohydrates (*e.g.*, hyaluronic acid and trehalose), aromatic compounds (*e.g.*, aromatic amines, vanillin, and indigo), vitamins and cofactors (as described in Ullmann's Encyclopedia of Industrial Chemistry, vol. A27, "Vitamins", p. 443-613 (1996) VCH: Weinheim and references therein; and Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research – Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press, (1995)), enzymes, polyketides (Cane *et al.* (1998) *Science* 282: 63-68), and all other chemicals described in Gutcho (1983) Chemicals by Fermentation, Noyes Data Corporation, ISBN: 0818805086 and references therein. The metabolism and uses of certain of these fine chemicals are further explicated below.

### *A. Amino Acid Metabolism and Uses*

Amino acids comprise the basic structural units of all proteins, and as such are essential for normal cellular functioning in all organisms. The term "amino acid" is art-recognized. The proteinogenic amino acids, of which there are 20 species, serve as structural units for proteins, in which they are linked by peptide bonds, while the nonproteinogenic amino acids (hundreds of which are known) are not normally found in proteins (see Ullmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97 VCH: Weinheim (1985)). Amino acids may be in the D- or L- optical configuration, though L-amino acids are generally the only type found in naturally-occurring proteins. Biosynthetic and degradative pathways of each of the 20 proteinogenic amino acids

have been well characterized in both prokaryotic and eukaryotic cells (see, for example, Stryer, L. Biochemistry, 3<sup>rd</sup> edition, pages 578-590 (1988)). The 'essential' amino acids (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine), so named because they are generally a nutritional requirement due to the complexity of their biosyntheses, are readily converted by simple biosynthetic pathways to the remaining 11 'nonessential' amino acids (alanine, arginine, asparagine, aspartate, cysteine, glutamate, glutamine, glycine, proline, serine, and tyrosine). Higher animals do retain the ability to synthesize some of these amino acids, but the essential amino acids must be supplied from the diet in order for normal protein synthesis to occur.

Aside from their function in protein biosynthesis, these amino acids are interesting chemicals in their own right, and many have been found to have various applications in the food, feed, chemical, cosmetics, agriculture, and pharmaceutical industries. Lysine is an important amino acid in the nutrition not only of humans, but also of monogastric animals such as poultry and swine. Glutamate is most commonly used as a flavor additive (mono-sodium glutamate, MSG) and is widely used throughout the food industry, as are aspartate, phenylalanine, glycine, and cysteine. Glycine, L-methionine and tryptophan are all utilized in the pharmaceutical industry. Glutamine, valine, leucine, isoleucine, histidine, arginine, proline, serine and alanine are of use in both the pharmaceutical and cosmetics industries. Threonine, tryptophan, and D/ L-methionine are common feed additives. (Leuchtenberger, W. (1996) Amino acids – technical production and use, p. 466-502 in Rehm *et al.* (eds.) Biotechnology vol. 6, chapter 14a, VCH: Weinheim). Additionally, these amino acids have been found to be useful as precursors for the synthesis of synthetic amino acids and proteins, such as N-acetylcysteine, S-carboxymethyl-L-cysteine, (S)-5-hydroxytryptophan, and others described in Ulmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97, VCH: Weinheim, 1985.

The biosynthesis of these natural amino acids in organisms capable of producing them, such as bacteria, has been well characterized (for review of bacterial amino acid biosynthesis and regulation thereof, see Umbarger, H.E. (1978) *Ann. Rev. Biochem.* 47: 533-606). Glutamate is synthesized by the reductive amination of  $\alpha$ -ketoglutarate, an intermediate in the citric acid cycle. Glutamine, proline, and arginine are each subsequently produced from glutamate. The biosynthesis of serine is a three-

step process beginning with 3-phosphoglycerate (an intermediate in glycolysis), and resulting in this amino acid after oxidation, transamination, and hydrolysis steps. Both cysteine and glycine are produced from serine; the former by the condensation of homocysteine with serine, and the latter by the transferal of the side-chain  $\beta$ -carbon atom to tetrahydrofolate, in a reaction catalyzed by serine transhydroxymethylase. Phenylalanine, and tyrosine are synthesized from the glycolytic and pentose phosphate pathway precursors erythrose 4-phosphate and phosphoenolpyruvate in a 9-step biosynthetic pathway that differ only at the final two steps after synthesis of prephenate. Tryptophan is also produced from these two initial molecules, but its synthesis is an 11-step pathway. Tyrosine may also be synthesized from phenylalanine, in a reaction catalyzed by phenylalanine hydroxylase. Alanine, valine, and leucine are all biosynthetic products of pyruvate, the final product of glycolysis. Aspartate is formed from oxaloacetate, an intermediate of the citric acid cycle. Asparagine, methionine, threonine, and lysine are each produced by the conversion of aspartate. Isoleucine is formed from threonine. A complex 9-step pathway results in the production of histidine from 5-phosphoribosyl-1-pyrophosphate, an activated sugar.

Amino acids in excess of the protein synthesis needs of the cell cannot be stored, and are instead degraded to provide intermediates for the major metabolic pathways of the cell (for review see Stryer, L. Biochemistry 3<sup>rd</sup> ed. Ch. 21 "Amino Acid Degradation and the Urea Cycle" p. 495-516 (1988)). Although the cell is able to convert unwanted amino acids into useful metabolic intermediates, amino acid production is costly in terms of energy, precursor molecules, and the enzymes necessary to synthesize them. Thus it is not surprising that amino acid biosynthesis is regulated by feedback inhibition, in which the presence of a particular amino acid serves to slow or entirely stop its own production (for overview of feedback mechanisms in amino acid biosynthetic pathways, see Stryer, L. Biochemistry, 3<sup>rd</sup> ed. Ch. 24: "Biosynthesis of Amino Acids and Heme" p. 575-600 (1988)). Thus, the output of any particular amino acid is limited by the amount of that amino acid present in the cell.

### 30 B. Vitamin, Cofactor, and Nutraceutical Metabolism and Uses

Vitamins, cofactors, and nutraceuticals comprise another group of molecules which the higher animals have lost the ability to synthesize and so must ingest, although

they are readily synthesized by other organisms, such as bacteria. These molecules are either bioactive substances themselves, or are precursors of biologically active substances which may serve as electron carriers or intermediates in a variety of metabolic pathways. Aside from their nutritive value, these compounds also have significant industrial value as coloring agents, antioxidants, and catalysts or other processing aids. (For an overview of the structure, activity, and industrial applications of these compounds, see, for example, Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996.) The term "vitamin" is art-recognized, and includes nutrients which are required by an organism for normal functioning, but which that organism cannot synthesize by itself. The group of vitamins may encompass cofactors and nutraceutical compounds. The language "cofactor" includes nonproteinaceous compounds required for a normal enzymatic activity to occur. Such compounds may be organic or inorganic; the cofactor molecules of the invention are preferably organic. The term "nutraceutical" includes dietary supplements having health benefits in plants and animals, particularly humans. Examples of such molecules are vitamins, antioxidants, and also certain lipids (*e.g.*, polyunsaturated fatty acids).

The biosynthesis of these molecules in organisms capable of producing them, such as bacteria, has been largely characterized (Ullman's Encyclopedia of Industrial Chemistry, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley & Sons; Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research – Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press: Champaign, IL X, 374 S).

Thiamin (vitamin B<sub>1</sub>) is produced by the chemical coupling of pyrimidine and thiazole moieties. Riboflavin (vitamin B<sub>2</sub>) is synthesized from guanosine-5'-triphosphate (GTP) and ribose-5'-phosphate. Riboflavin, in turn, is utilized for the synthesis of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). The family of compounds collectively termed 'vitamin B<sub>6</sub>' (*e.g.*, pyridoxine, pyridoxamine, pyridoxa-5'-phosphate, and the commercially used pyridoxin hydrochloride) are all derivatives of the common structural unit, 5-hydroxy-6-methylpyridine. Pantothenate (pantothenic

acid, (R)-(+)-N-(2,4-dihydroxy-3,3-dimethyl-1-oxobutyl)- $\beta$ -alanine) can be produced either by chemical synthesis or by fermentation. The final steps in pantothenate biosynthesis consist of the ATP-driven condensation of  $\beta$ -alanine and pantoic acid. The enzymes responsible for the biosynthesis steps for the conversion to pantoic acid, to  $\beta$ -alanine and for the condensation to panthothenic acid are known. The metabolically active form of pantothenate is Coenzyme A, for which the biosynthesis proceeds in 5 enzymatic steps. Pantothenate, pyridoxal-5'-phosphate, cysteine and ATP are the precursors of Coenzyme A. These enzymes not only catalyze the formation of panthothante, but also the production of (R)-pantoic acid, (R)-pantolacton, (R)-panthenol (provitamin B<sub>5</sub>), pantetheine (and its derivatives) and coenzyme A.

Biotin biosynthesis from the precursor molecule pimeloyl-CoA in microorganisms has been studied in detail and several of the genes involved have been identified. Many of the corresponding proteins have been found to also be involved in Fe-cluster synthesis and are members of the nifS class of proteins. Lipoic acid is derived from octanoic acid, and serves as a coenzyme in energy metabolism, where it becomes part of the pyruvate dehydrogenase complex and the  $\alpha$ -ketoglutarate dehydrogenase complex. The folates are a group of substances which are all derivatives of folic acid, which in turn is derived from L-glutamic acid, p-amino-benzoic acid and 6-methylpterin. The biosynthesis of folic acid and its derivatives, starting from the metabolism intermediates guanosine-5'-triphosphate (GTP), L-glutamic acid and p-amino-benzoic acid has been studied in detail in certain microorganisms.

Corrinoids (such as the cobalamines and particularly vitamin B<sub>12</sub>) and porphyrines belong to a group of chemicals characterized by a tetrapyrrole ring system. The biosynthesis of vitamin B<sub>12</sub> is sufficiently complex that it has not yet been completely characterized, but many of the enzymes and substrates involved are now known. Nicotinic acid (nicotinate), and nicotinamide are pyridine derivatives which are also termed 'niacin'. Niacin is the precursor of the important coenzymes NAD (nicotinamide adenine dinucleotide) and NADP (nicotinamide adenine dinucleotide phosphate) and their reduced forms.

The large-scale production of these compounds has largely relied on cell-free chemical syntheses, though some of these chemicals have also been produced by large-scale culture of microorganisms, such as riboflavin, Vitamin B<sub>6</sub>, pantothenate, and



biotin. Only Vitamin B<sub>12</sub> is produced solely by fermentation, due to the complexity of its synthesis. *In vitro* methodologies require significant inputs of materials and time, often at great cost.

5    C. *Purine, Pyrimidine, Nucleoside and Nucleotide Metabolism and Uses*

Purine and pyrimidine metabolism genes and their corresponding proteins are important targets for the therapy of tumor diseases and viral infections. The language "purine" or "pyrimidine" includes the nitrogenous bases which are constituents of nucleic acids, co-enzymes, and nucleotides. The term "nucleotide" includes the basic structural units of nucleic acid molecules, which are comprised of a nitrogenous base, a pentose sugar (in the case of RNA, the sugar is ribose; in the case of DNA, the sugar is D-deoxyribose), and phosphoric acid. The language "nucleoside" includes molecules which serve as precursors to nucleotides, but which are lacking the phosphoric acid moiety that nucleotides possess. By inhibiting the biosynthesis of these molecules, or their mobilization to form nucleic acid molecules, it is possible to inhibit RNA and DNA synthesis; by inhibiting this activity in a fashion targeted to cancerous cells, the ability of tumor cells to divide and replicate may be inhibited. Additionally, there are nucleotides which do not form nucleic acid molecules, but rather serve as energy stores (*i.e.*, AMP) or as coenzymes (*i.e.*, FAD and NAD).

20       Several publications have described the use of these chemicals for these medical indications, by influencing purine and/or pyrimidine metabolism (*e.g.* Christopherson, R.I. and Lyons, S.D. (1990) "Potent inhibitors of *de novo* pyrimidine and purine biosynthesis as chemotherapeutic agents." *Med. Res. Reviews* 10: 505-548). Studies of enzymes involved in purine and pyrimidine metabolism have been focused on the development of new drugs which can be used, for example, as immunosuppressants or anti-proliferants (Smith, J.L., (1995) "Enzymes in nucleotide synthesis." *Curr. Opin. Struct. Biol.* 5: 752-757; (1995) *Biochem Soc. Transact.* 23: 877-902). However, purine and pyrimidine bases, nucleosides and nucleotides have other utilities: as intermediates in the biosynthesis of several fine chemicals (*e.g.*, thiamine, S-adenosyl-methionine, folates, or riboflavin), as energy carriers for the cell (*e.g.*, ATP or GTP), and for chemicals themselves, commonly used as flavor enhancers (*e.g.*, IMP or GMP) or for several medicinal applications (see, for example, Kuninaka, A. (1996) Nucleotides and



Related Compounds in Biotechnology vol. 6, Rehm *et al.*, eds. VCH: Weinheim, p. 561-612). Also, enzymes involved in purine, pyrimidine, nucleoside, or nucleotide metabolism are increasingly serving as targets against which chemicals for crop protection, including fungicides, herbicides and insecticides, are developed.

- 5           The metabolism of these compounds in bacteria has been characterized (for reviews see, for example, Zalkin, H. and Dixon, J.E. (1992) "*de novo* purine nucleotide biosynthesis", in: Progress in Nucleic Acid Research and Molecular Biology, vol. 42, Academic Press:, p. 259-287; and Michal, G. (1999) "Nucleotides and Nucleosides", Chapter 8 in: Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology,
- 10   Wiley: New York). Purine metabolism has been the subject of intensive research, and is essential to the normal functioning of the cell. Impaired purine metabolism in higher animals can cause severe disease, such as gout. Purine nucleotides are synthesized from ribose-5-phosphate, in a series of steps through the intermediate compound inosine-5'-phosphate (IMP), resulting in the production of guanosine-5'-monophosphate (GMP) or
- 15   adenosine-5'-monophosphate (AMP), from which the triphosphate forms utilized as nucleotides are readily formed. These compounds are also utilized as energy stores, so their degradation provides energy for many different biochemical processes in the cell. Pyrimidine biosynthesis proceeds by the formation of uridine-5'-monophosphate (UMP) from ribose-5-phosphate. UMP, in turn, is converted to cytidine-5'-triphosphate (CTP).
- 20   The deoxy- forms of all of these nucleotides are produced in a one step reduction reaction from the diphosphate ribose form of the nucleotide to the diphosphate deoxyribose form of the nucleotide. Upon phosphorylation, these molecules are able to participate in DNA synthesis.

#### 25   D. Trehalose Metabolism and Uses

- Trehalose consists of two glucose molecules, bound in  $\alpha$ ,  $\alpha$ -1,1 linkage. It is commonly used in the food industry as a sweetener, an additive for dried or frozen foods, and in beverages. However, it also has applications in the pharmaceutical, cosmetics and biotechnology industries (see, for example, Nishimoto *et al.*, (1998) U.S.
- 30   Patent No. 5,759,610; Singer, M.A. and Lindquist, S. (1998) *Trends Biotech.* 16: 460-467; Paiva, C.L.A. and Panek, A.D. (1996) *Biotech. Ann. Rev.* 2: 293-314; and Shiosaka, M. (1997) *J. Japan* 172: 97-102). Trehalose is produced by enzymes from

many microorganisms and is naturally released into the surrounding medium, from which it can be collected using methods known in the art.

## II. Elements and Methods of the Invention

5           The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as MP nucleic acid and protein molecules, which play a role in or function in one or more cellular metabolic pathways. In one embodiment, the MP molecules catalyze an enzymatic reaction involving one or more amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic  
10 pathways. In a preferred embodiment, the activity of the MP molecules of the present invention in one or more *C. glutamicum* metabolic pathways for amino acids, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides or trehalose has an impact on the production of a desired fine chemical by this organism. In a particularly preferred embodiment, the MP molecules of the invention are modulated in activity, such that the  
15 *C. glutamicum* metabolic pathways in which the MP proteins of the invention are involved are modulated in efficiency or output, which either directly or indirectly modulates the production or efficiency of production of a desired fine chemical by *C. glutamicum*.

          The language, "MP protein" or "MP polypeptide" includes proteins which play  
20 a role in, *e.g.*, catalyze an enzymatic reaction, in one or more amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside or trehalose metabolic pathways. Examples of MP proteins include those encoded by the MP genes set forth in Table 1 and by the odd-numbered SEQ ID NOs. The terms "MP gene" or "MP nucleic acid sequence" include nucleic acid sequences encoding an MP protein, which consist of a  
25 coding region and also corresponding untranslated 5' and 3' sequence regions. Examples of MP genes include those set forth in Table 1. The terms "production" or "productivity" are art-recognized and include the concentration of the fermentation product (for example, the desired fine chemical) formed within a given time and a given fermentation volume (*e.g.*, kg product per hour per liter). The term "efficiency of  
30 production" includes the time required for a particular level of production to be achieved (for example, how long it takes for the cell to attain a particular rate of output of a fine chemical). The term "yield" or "product/carbon yield" is art-recognized and includes

the efficiency of the conversion of the carbon source into the product (*i.e.*, fine chemical). This is generally written as, for example, kg product per kg carbon source. By increasing the yield or production of the compound, the quantity of recovered molecules, or of useful recovered molecules of that compound in a given amount of culture over a given amount of time is increased. The terms "biosynthesis" or a "biosynthetic pathway" are art-recognized and include the synthesis of a compound, preferably an organic compound, by a cell from intermediate compounds in what may be a multistep and highly regulated process. The terms "degradation" or a "degradation pathway" are art-recognized and include the breakdown of a compound, preferably an organic compound, by a cell to degradation products (generally speaking, smaller or less complex molecules) in what may be a multistep and highly regulated process. The language "metabolism" is art-recognized and includes the totality of the biochemical reactions that take place in an organism. The metabolism of a particular compound, then, (*e.g.*, the metabolism of an amino acid such as glycine) comprises the overall biosynthetic, modification, and degradation pathways in the cell related to this compound.

In another embodiment, the MP molecules of the invention are capable of modulating the production of a desired molecule, such as a fine chemical, in a microorganism such as *C. glutamicum*. Using recombinant genetic techniques, one or more of the biosynthetic or degradative enzymes of the invention for amino acids, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, or trehalose may be manipulated such that its function is modulated. For example, a biosynthetic enzyme may be improved in efficiency, or its allosteric control region destroyed such that feedback inhibition of production of the compound is prevented. Similarly, a degradative enzyme may be deleted or modified by substitution, deletion, or addition such that its degradative activity is lessened for the desired compound without impairing the viability of the cell. In each case, the overall yield or rate of production of one of these desired fine chemicals may be increased.

It is also possible that such alterations in the protein and nucleotide molecules of the invention may improve the production of other fine chemicals besides the amino acids, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, and trehalose. Metabolism of any one compound is necessarily intertwined with other biosynthetic and

degradative pathways within the cell, and necessary cofactors, intermediates, or substrates in one pathway are likely supplied or limited by another such pathway.

Therefore, by modulating the activity of one or more of the proteins of the invention, the production or efficiency of activity of another fine chemical biosynthetic or degradative

5 pathway may be impacted. For example, amino acids serve as the structural units of all proteins, yet may be present intracellularly in levels which are limiting for protein synthesis; therefore, by increasing the efficiency of production or the yields of one or more amino acids within the cell, proteins, such as biosynthetic or degradative proteins, may be more readily synthesized. Likewise, an alteration in a metabolic pathway  
10 enzyme such that a particular side reaction becomes more or less favored may result in the over- or under-production of one or more compounds which are utilized as intermediates or substrates for the production of a desired fine chemical.

The isolated nucleic acid sequences of the invention are contained within the genome of a *Corynebacterium glutamicum* strain available through the American Type  
15 Culture Collection, given designation ATCC 13032. The nucleotide sequence of the isolated *C. glutamicum* MP DNAs and the predicted amino acid sequences of the *C. glutamicum* MP proteins are shown in the Sequence Listing as odd-numbered SEQ ID NOs and even-numbered SEQ ID NOs, respectively. Computational analyses were performed which classified and/or identified these nucleotide sequences as  
20 sequences which encode metabolic pathway proteins.

The present invention also pertains to proteins which have an amino acid sequence which is substantially homologous to an amino acid sequence of the invention (e.g., the sequence of an even-numbered SEQ ID NO of the Sequence Listing). As used herein, a protein which has an amino acid sequence which is substantially homologous  
25 to a selected amino acid sequence is least about 50% homologous to the selected amino acid sequence, e.g., the entire selected amino acid sequence. A protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence can also be least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, or 90-95%, and most preferably at least about  
30 96%, 97%, 98%, 99% or more homologous to the selected amino acid sequence.

The MP protein or a biologically active portion or fragment thereof of the invention can catalyze an enzymatic reaction in one or more amino acid, vitamin,

cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathways, or have one or more of the activities set forth in Table 1.

Various aspects of the invention are described in further detail in the following subsections:

5

#### *A. Isolated Nucleic Acid Molecules*

One aspect of the invention pertains to isolated nucleic acid molecules that encode MP polypeptides or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes or primers for the identification or  
10 amplification of MP-encoding nucleic acid (*e.g.*, MP DNA). As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. This term also encompasses untranslated sequence located at both the 3' and 5' ends of the coding region of the gene: at least about 100 nucleotides  
15 of sequence upstream from the 5' end of the coding region and at least about 20 nucleotides of sequence downstream from the 3' end of the coding region of the gene. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic  
20 acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated MP nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank  
25 the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived (*e.g.*, a *C. glutamicum* cell). Moreover, an "isolated" nucleic acid molecule, such as a DNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized.

30

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having a nucleotide sequence of an odd-numbered SEQ ID NO of the Sequence Listing, or a portion thereof, can be isolated using standard molecular biology techniques and the



sequence information provided herein. For example, a *C. glutamicum* MP DNA can be isolated from a *C. glutamicum* library using all or portion of one of the odd-numbered SEQ ID NO sequences of the Sequence Listing as a hybridization probe and standard hybridization techniques (*e.g.*, as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Moreover, a nucleic acid molecule encompassing all or a portion of one of the nucleic acid sequences of the invention (*e.g.*, an odd-numbered SEQ ID NO:) can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this sequence (*e.g.*, a nucleic acid molecule encompassing all or a portion of one of the nucleic acid sequences of the invention (*e.g.*, an odd-numbered SEQ ID NO of the Sequence Listing) can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this same sequence). For example, mRNA can be isolated from normal endothelial cells (*e.g.*, by the guanidinium-thiocyanate extraction procedure of Chirgwin *et al.* (1979) *Biochemistry* 18: 5294-5299) and DNA can be prepared using reverse transcriptase (*e.g.*, Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, MD; or AMV reverse transcriptase, available from Seikagaku America, Inc., St. Petersburg, FL). Synthetic oligonucleotide primers for polymerase chain reaction amplification can be designed based upon one of the nucleotide sequences shown in the Sequence Listing. A nucleic acid of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to an MP nucleotide sequence can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises one of the nucleotide sequences shown in the Sequence Listing. The nucleic acid sequences of the invention, as set forth in the Sequence Listing, correspond to the *Corynebacterium glutamicum* MP DNAs of the invention. This DNA comprises sequences encoding MP proteins (*i.e.*, the "coding region", indicated in each odd-numbered SEQ ID NO: sequence in the Sequence Listing), as well as 5' untranslated



sequences and 3' untranslated sequences, also indicated in each odd-numbered SEQ ID NO: in the Sequence Listing. Alternatively, the nucleic acid molecule can comprise only the coding region of any of the nucleic acid sequences of the Sequence Listing.

For the purposes of this application, it will be understood that each of the nucleic acid and amino acid sequences set forth in the Sequence Listing has an identifying RXA, RXN, RXS, or RXC number having the designation "RXA", "RXN", "RXS", or "RXC" followed by 5 digits (*i.e.*, RXA00007, RXN00023, RXS00116, or RXC00128). Each of the nucleic acid sequences comprises up to three parts: a 5' upstream region, a coding region, and a downstream region. Each of these three regions is identified by the same RXA, RXN, RXS, or RXC designation to eliminate confusion. The recitation "one of the odd-numbered sequences of the Sequence Listing", then, refers to any of the nucleic acid sequences in the Sequence Listing, which may also be distinguished by their differing RXA, RXN, RXS, or RXC designations. The coding region of each of these sequences is translated into a corresponding amino acid sequence, which is also set forth in the Sequence Listing, as an even-numbered SEQ ID NO: immediately following the corresponding nucleic acid sequence. For example, the coding region for RXA02229 is set forth in SEQ ID NO:1, while the amino acid sequence which it encodes is set forth as SEQ ID NO:2. The sequences of the nucleic acid molecules of the invention are identified by the same RXA, RXN, RXS, or RXC designations as the amino acid molecules which they encode, such that they can be readily correlated. For example, the amino acid sequences designated RXA02229, RX00351, RXS02970, and RXC02390 are translations of the coding regions of the nucleotide sequences of nucleic acid molecules RXA02229, RX00351, RXS02970, and RXC02390, respectively. The correspondence between the RXA, RXN, RXS, and RXC nucleotide and amino acid sequences of the invention and their assigned SEQ ID NOs is set forth in Table 1.

Several of the genes of the invention are "F-designated genes". An F-designated gene includes those genes set forth in Table 1 which have an 'F' in front of the RXA, RXN, RXS, or RXC designation. For example, SEQ ID NO:5, designated, as indicated on Table 1, as "F RXA01009", is an F-designated gene, as are SEQ ID NOs: 73, 75, and 77 (designated on Table 1 as "F RXA00007", "F RXA00364", and "F RXA00367", respectively).

In one embodiment, the nucleic acid molecules of the present invention are not intended to include *C. glutamicum* those compiled in Table 2. In the case of the dapD gene, a sequence for this gene was published in Wehrmann, A., *et al.* (1998) *J. Bacteriol.* 180(12): 3159-3165. However, the sequence obtained by the inventors of the present application is significantly longer than the published version. It is believed that the published version relied on an incorrect start codon, and thus represents only a fragment of the actual coding region.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of one of the nucleotide sequences of the invention (*e.g.*, a sequence of an odd-numbered SEQ ID NO: of the Sequence Listing), or a portion thereof. A nucleic acid molecule which is complementary to one of the nucleotide sequences of the invention is one which is sufficiently complementary to one of the nucleotide sequences shown in the Sequence Listing (*e.g.*, the sequence of an odd-numbered SEQ ID NO:) such that it can hybridize to one of the nucleotide sequences of the invention, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence of the invention (*e.g.*, a sequence of an odd-numbered SEQ ID NO: of the Sequence Listing), or a portion thereof. Ranges and identity values intermediate to the above-recited ranges, (*e.g.*, 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are intended to be included. In an additional preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes, *e.g.*, hybridizes under stringent conditions, to one of the nucleotide sequences of the invention, or a portion thereof.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the coding region of the sequence of one of the odd-numbered SEQ ID NOs

of the Sequence Listing, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of an MP protein. The nucleotide sequences determined from the cloning of the MP genes from *C. glutamicum* allows for the generation of probes and primers designed for use in identifying and/or cloning MP  
5 homologues in other cell types and organisms, as well as MP homologues from other *Corynebacteria* or related species. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50 or 75 consecutive nucleotides of a sense strand of one  
10 of the nucleotide sequences of the invention (*e.g.*, a sequence of one of the odd-numbered SEQ ID NOs of the Sequence Listing), an anti-sense sequence of one of these sequences, or naturally occurring mutants thereof. Primers based on a nucleotide sequence of the invention can be used in PCR reactions to clone MP homologues. Probes based on the MP nucleotide sequences can be used to detect transcripts or  
15 genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, *e.g.* the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells which misexpress an MP protein, such as by measuring a level of an MP-encoding  
20 nucleic acid in a sample of cells from a subject *e.g.*, detecting MP mRNA levels or determining whether a genomic MP gene has been mutated or deleted.

In one embodiment, the nucleic acid molecule of the invention encodes a protein or portion thereof which includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of the invention (*e.g.*, a sequence of an even-  
25 numbered SEQ ID NO of the Sequence Listing) such that the protein or portion thereof maintains the ability to catalyze an enzymatic reaction in an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway. As used herein, the language "sufficiently homologous" refers to proteins or portions thereof which have amino acid sequences which include a minimum number of identical or  
30 equivalent (*e.g.*, an amino acid residue which has a similar side chain as an amino acid residue in a sequence of one of the even-numbered SEQ ID NOs of the Sequence Listing) amino acid residues to an amino acid sequence of the invention such that the

protein or portion thereof is able to catalyze an enzymatic reaction in a *C. glutamicum* amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside or trehalose metabolic pathway. Protein members of such metabolic pathways, as described herein, function to catalyze the biosynthesis or degradation of one or more of: amino acids, vitamins, cofactors, nutraceuticals, nucleotides, nucleosides, or trehalose. Examples of such activities are also described herein. Thus, "the function of an MP protein" contributes to the overall functioning of one or more such metabolic pathway and contributes, either directly or indirectly, to the yield, production, and/or efficiency of production of one or more fine chemicals. Examples of MP protein activities are set forth in Table 1.

In another embodiment, the protein is at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing).

Portions of proteins encoded by the MP nucleic acid molecules of the invention are preferably biologically active portions of one of the MP proteins. As used herein, the term "biologically active portion of an MP protein" is intended to include a portion, *e.g.*, a domain/motif, of an MP protein that catalyzes an enzymatic reaction in one or more *C. glutamicum* amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathways, or has an activity as set forth in Table 1. To determine whether an MP protein or a biologically active portion thereof can catalyze an enzymatic reaction in an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway, an assay of enzymatic activity may be performed. Such assay methods are well known to those of ordinary skill in the art, as detailed in Example 8 of the Exemplification.

Additional nucleic acid fragments encoding biologically active portions of an MP protein can be prepared by isolating a portion of one of the amino acid sequences of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing), expressing the encoded portion of the MP protein or peptide (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the MP protein or peptide.

The invention further encompasses nucleic acid molecules that differ from one of the nucleotide sequences of the invention (*e.g.*, a sequence of an odd-numbered SEQ ID NO: of the Sequence Listing) (and portions thereof) due to degeneracy of the genetic code and thus encode the same MP protein as that encoded by the nucleotide sequences of the invention. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in the Sequence Listing (*e.g.*, an even-numbered SEQ ID NO:). In a still further embodiment, the nucleic acid molecule of the invention encodes a full length *C. glutamicum* protein which is substantially homologous to an amino acid sequence of the invention (encoded by an open reading frame shown in an odd-numbered SEQ ID NO: of the Sequence Listing).

It will be understood by one of ordinary skill in the art that in one embodiment the sequences of the invention are not meant to include the sequences of the prior art, such as those Genbank sequences set forth in Tables 2 or 4 which were available prior to the present invention. In one embodiment, the invention includes nucleotide and amino acid sequences having a percent identity to a nucleotide or amino acid sequence of the invention which is greater than that of a sequence of the prior art (*e.g.*, a Genbank sequence (or the protein encoded by such a sequence) set forth in Tables 2 or 4). For example, the invention includes a nucleotide sequence which is greater than and/or at least 40% identical to the nucleotide sequence designated RXA00115 (SEQ ID NO:185), a nucleotide sequence which is greater than and/or at least % identical to the nucleotide sequence designated RXA00131 (SEQ ID NO:991), and a nucleotide sequence which is greater than and/or at least 39% identical to the nucleotide sequence designated RXA00219 (SEQ ID NO:345). One of ordinary skill in the art would be able to calculate the lower threshold of percent identity for any given sequence of the invention by examining the GAP-calculated percent identity scores set forth in Table 4 for each of the three top hits for the given sequence, and by subtracting the highest GAP-calculated percent identity from 100 percent. One of ordinary skill in the art will also appreciate that nucleic acid and amino acid sequences having percent identities greater than the lower threshold so calculated (*e.g.*, at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%,



74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more identical) are also encompassed by the invention.

In addition to the *C. glutamicum* MP nucleotide sequences set forth in the Sequence Listing as odd-numbered SEQ ID NOs, it will be appreciated by one of ordinary skill in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of MP proteins may exist within a population (*e.g.*, the *C. glutamicum* population). Such genetic polymorphism in the MP gene may exist among individuals within a population due to natural variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding an MP protein, preferably a *C. glutamicum* MP protein. Such natural variations can typically result in 1-5% variance in the nucleotide sequence of the MP gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in MP that are the result of natural variation and that do not alter the functional activity of MP proteins are intended to be within the scope of the invention.

Nucleic acid molecules corresponding to natural variants and non-*C. glutamicum* homologues of the *C. glutamicum* MP DNA of the invention can be isolated based on their homology to the *C. glutamicum* MP nucleic acid disclosed herein using the *C. glutamicum* DNA, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising a nucleotide sequence of an odd-numbered SEQ ID NO: of the Sequence Listing. In other embodiments, the nucleic acid is at least 30, 50, 100, 250 or more nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65%, more preferably at least about 70%, and even more preferably at least about 75% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to one of ordinary skill in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6.



A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a nucleotide sequence of the invention corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein). In one embodiment, the nucleic acid encodes a natural *C. glutamicum* MP protein.

10 In addition to naturally-occurring variants of the MP sequence that may exist in the population, one of ordinary skill in the art will further appreciate that changes can be introduced by mutation into a nucleotide sequence of the invention, thereby leading to changes in the amino acid sequence of the encoded MP protein, without altering the functional ability of the MP protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in a nucleotide sequence of the invention. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of one of the MP proteins (*e.g.*, an even-numbered SEQ ID NO: of the Sequence Listing) without altering the activity of said MP protein, whereas an "essential" amino acid residue is required for MP protein activity.

15 20 Other amino acid residues, however, (*e.g.*, those that are not conserved or only semi-conserved in the domain having MP activity) may not be essential for activity and thus are likely to be amenable to alteration without altering MP activity.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding MP proteins that contain changes in amino acid residues that are not essential for MP activity. Such MP proteins differ in amino acid sequence from a sequence of an even-numbered SEQ ID NO: of the Sequence Listing yet retain at least one of the MP activities described herein. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50% homologous to an amino acid sequence of the invention and is capable of catalyzing an enzymatic reaction in an amino acid, vitamin, cofactor, nutraceutical, nucleotide, nucleoside, or trehalose metabolic pathway, or has one or more activities set forth in Table 1. Preferably, the protein encoded by the nucleic

25 30

acid molecule is at least about 50-60% homologous to the amino acid sequence of one of the odd-numbered SEQ ID NOs of the Sequence Listing, more preferably at least about 60-70% homologous to one of these sequences, even more preferably at least about 70-80%, 80-90%, 90-95% homologous to one of these sequences, and most preferably at least about 96%, 97%, 98%, or 99% homologous to one of the amino acid sequences of the invention.

To determine the percent homology of two amino acid sequences (*e.g.*, one of the amino acid sequences of the invention and a mutant form thereof) or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in one sequence (*e.g.*, one of the amino acid sequences of the invention) is occupied by the same amino acid residue or nucleotide as the corresponding position in the other sequence (*e.g.*, a mutant form of the amino acid sequence), then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % homology = # of identical positions/total # of positions x 100).

An isolated nucleic acid molecule encoding an MP protein homologous to a protein sequence of the invention (*e.g.*, a sequence of an even-numbered SEQ ID NO: of the Sequence Listing) can be created by introducing one or more nucleotide substitutions, additions or deletions into a nucleotide sequence of the invention such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into one of the nucleotide sequences of the invention by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic

acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine,

5 phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in an MP protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an MP coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for an MP activity described  
10 herein to identify mutants that retain MP activity. Following mutagenesis of the nucleotide sequence of one of the odd-numbered SEQ ID NOs of the Sequence Listing, the encoded protein can be expressed recombinantly and the activity of the protein can be determined using, for example, assays described herein (see Example 8 of the Exemplification).

15 In addition to the nucleic acid molecules encoding MP proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded DNA molecule or complementary to an mRNA  
20 sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire MP coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an MP protein. The term "coding region" refers to the region of the nucleotide sequence  
25 comprising codons which are translated into amino acid residues (*e.g.*, the entire coding region of SEQ ID NO. 1 (RXA02229) comprises nucleotides 1 to 825). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding MP. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into  
30 amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding MP disclosed herein (*e.g.*, the sequences set forth as odd-numbered SEQ ID NOs in the Sequence Listing), antisense

nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of MP mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of MP mRNA. For  
5 example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of MP mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an  
10 antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified  
15 nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine,  
20 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid  
25 methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from  
30 the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a cell or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an MCT protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by  
5 conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. The antisense molecule can be modified such that it specifically binds to a receptor or an antigen expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecule to a peptide or  
10 an antibody which binds to a cell surface receptor or antigen. The antisense nucleic acid molecule can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong prokaryotic, viral, or eukaryotic promoter are preferred.

15 In yet another embodiment, the antisense nucleic acid molecule of the invention is an  $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-  
20 methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they  
25 have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave MP mRNA transcripts to thereby inhibit translation of MP mRNA. A ribozyme having specificity for an MP-encoding nucleic acid can be designed based upon the nucleotide sequence of an MP DNA disclosed herein (*i.e.*, SEQ ID NO: 1  
30 (RXA02229). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an MP-encoding mRNA. See, *e.g.*, Cech *et al.*



U.S. Patent No. 4,987,071 and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, MP mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

- 5           Alternatively, MP gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of an MP nucleotide sequence (*e.g.*, an MP promoter and/or enhancers) to form triple helical structures that prevent transcription of an MP gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and  
10   Maher, L.J. (1992) *Bioassays* 14(12):807-15.

#### *B. Recombinant Expression Vectors and Host Cells*

- Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an MP protein (or a portion thereof). As  
15   used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of  
20   autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to  
25   which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors,  
30   such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.